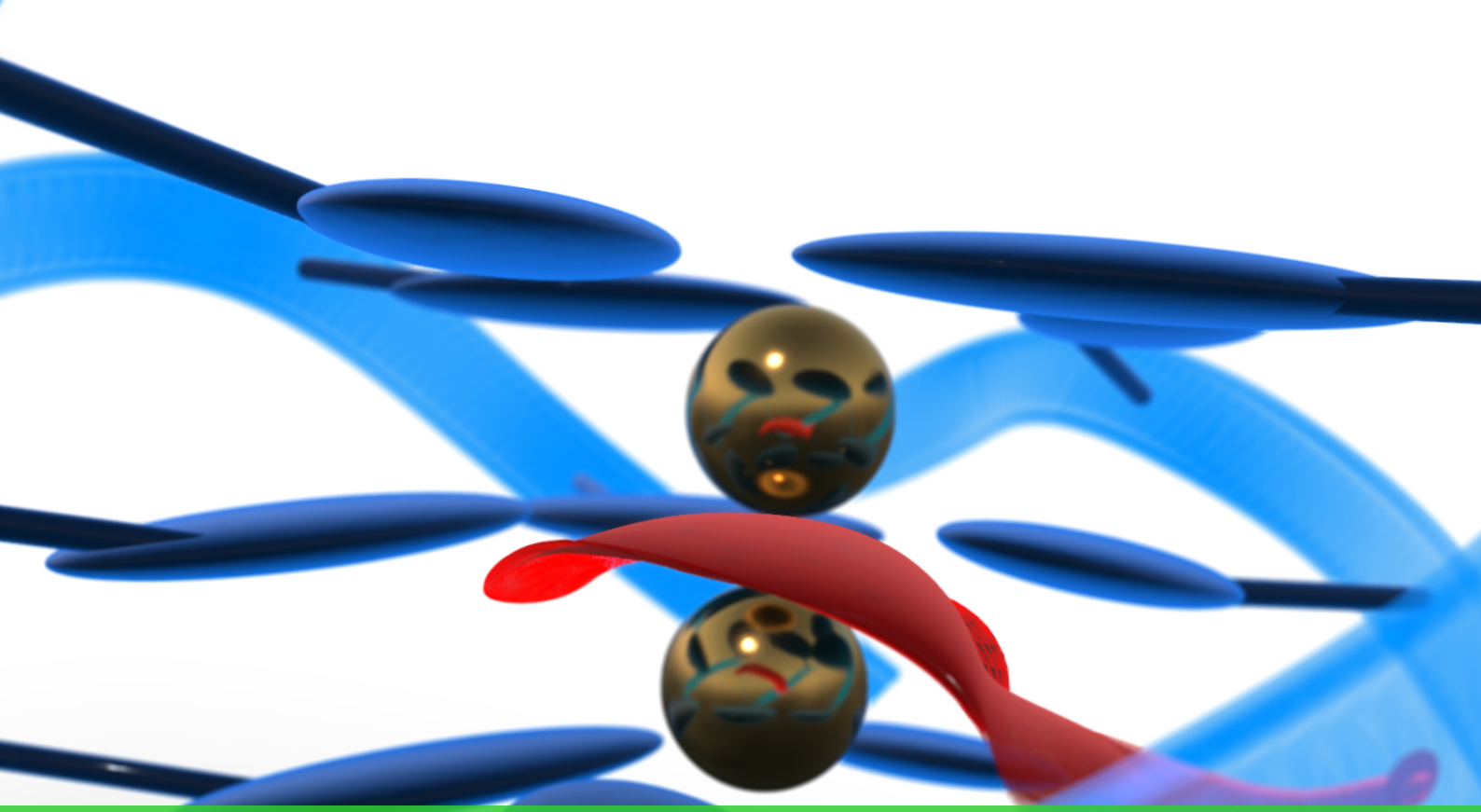


Modeling the influence of DNA lesion on the regulation of gene expression

Philosophiæ Doctor Thesis

Tom MICLOT

UNIVERSITÉ DE LORRAINE

UNIVERSITÀ DEGLI STUDI DI PALERMO

Philosophiæ Doctor Thesis

This research was performed in a Italian-French joint project between Università degli Studi di Palermo (Italia) and Université de Lorraine (France), founded by Università degli Studi di Palermo from November 2019 to October 2022. The project was co-supervised by Pr. Dr. Giampaolo Barone (Italia) and Pr. Dr. Antonio Monari (France).

Cover illustration – Artist's view of a G-quadruplex DNA with one damaged nucleotide (red), by Tom Miclot, based on the PDB:1KF1 structure.

First release in November 15, 2022

UNIVERSITÀ DEGLI STUDI DI PALERMO

UNIVERSITÉ DE LORRAINE

AREA QUALITÀ, PROGRAMMAZIONE E SUPPORTO STRATEGICO
SETTORE STRATEGIA PER LA RICERCA

U. O. DOTTORATI

Scienze Molecolari e Biomolecolari

Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche

CHIM/03

Tesi in cotutela con l'Università della Lorena (Nancy, Francia)

Modeling the influence of DNA lesion on the regulation of gene expression

IL DOTTORE

Tom MICLOT

IL COORDINATORE

Giovanna PITARRESI

IL TUTOR

Antonio MONARI

IL TUTOR

Giampaolo BARONE

CICLO XXXV

ANNO CONSEGUIMENTO TITOLO 2022

Modeling the influence of DNA lesion on the regulation of gene expression

by **Tom Miclot**

PhD thesis publicly defended on December 2022 in order to obtain the PhD degree in Chemistry, for the *Université de Lorraine*, and the PhD degree in Molecular and Biomolecular Sciences, for the *Università degli studi di Palermo*

Jury members

Rapporteur	Dr. Matteo Lambrughi	Danish Cancer Society Research Center
Rapporteur	Prof. Iñaki Tuñón	Universidad de Valencia
Examiner	A. Prof. Adriana Pietropaolo	Università Magna Graecia di Catanzaro
Examiner	Dr. Antonio Santoro	Università degli Studi di Messina
Examiner	Prof. Stéphanie Grandemange	Université de Lorraine
Director	Prof. Giampaolo Barone	Università degli Studi di Palermo
Co-Director	Prof. Antonio Monari	Université Paris Cité

Dottorati di Ricerca, UniPa and *Ecole doctorale C2MP*

Whatever your life is, make it useful. To do this, strive to make some progress, however modest it may seem to you, in the work you have chosen, and be certain that you will thus help the general welfare and increase the common heritage.

Gustave Eiffel

Personal translation

Acknowledgements

Child of the sun,
Your destiny is unparalleled,
Adventure is calling you,
Don't wait and run towards it.

Les Mystérieuses Cités d'or

With these few words, I would like to express my thanks to Giampaolo Barone, great baron of tiramissu, and Antonio Monari, the great connoisseur of the Chimay Red®. Thank you for being such good PhD directors. And I especially want to thank this beer they drank together and that gave birth to this PhD thesis project: *The Beerborn PhD*.

Also, I would like to thank all rapporteurs, Iñaki Tuñón, Matteo Lambrughi, for doing me the honor of accepting to examine my work.

Thanks to all my colleagues of both French and Italian laboratories, for their advice and their welcome during my internship and my thesis.

A big thank you to all my family, without whom I would not be here today. So it is important for me to thank you here for always supporting me and believing in me. I also thank my godfather and my godmother who were always present.

I also thank my friend Mathieu, known as Raven, who always has his head in the stars.

Contents

Curriculum vitae

Name	Tom Martin Manfred Miclot
Date of birth	18^{th} August 1994
Place of birth	Strasbourg, France
Nationality	French

Professional networks

Google Scholar *ResearchGate* *ORCID* *LinkedIn* *Scopus* *GitHub*

Diplomas

2020 Master Epistemology, History of Science and Technology
Université de Lorraine, Nancy, France

2019 Master in Biotechnologies
Université de Lorraine, Vandoeuvre-lès-Nancy, France

2017 Bachelor of Science in Life Science
Université de Lorraine, Vandoeuvre-lès-Nancy, France

2015 Advanced Technician's Certificate in Biotechnologies
Lycée Polyvalent Régional Stanislas, Villers-lès-Nancy, France

2013 Baccalaureate in Laboratory Science and Technology
Lycée Louis-Vincent, Metz, France

Laboratory experiences before PhD

2020 Master thesis - 6 month

Title : *Correlation between physical time and computer time.*

Supervisor : Dr. Baptiste Mélès

Place : Archives Henri Poincaré ; Nancy ; France

2019 Master thesis - 6 month

Title : *Cloning, expression, purification, crystallisation and structural resolution of armadillo repeat proteins fused to a target peptide and a chaperone protein for crystallisation.*

Supervisor : PD Dr. Peer Mittl

Place : Laboratory of Prof. Dr. Andreas Plückthun ; Biochemisches Institut ; Universität Zürich ; Switzerland

2018 Internship - 7 weeks

Title : *Study of the interaction between the damaged BRCA2 promoter (abasic site, thymine dimer and 64PP) and NF-kB by molecular dynamics.*

Supervisor : A.Pr Dr. Antonio Monari

Place : Laboratoire de Physique et Chimie Théoriques UMR7019 ; Université de Lorraine ; France

2017 Bachelor thesis - 7 weeks

Title : *Evaluation of the RNA content of pole-laying fish eggs of variable quality, by RT-qPCR.*

Supervisor : Dr. Bérénice Schaerlinger

Place : UR AFPA ; Université de Lorraine ; Nancy ;France

2015 Advanced technician certification thesis - 9 weeks

Title : *Extraction, purification and crystallization of the SCPb protease of* Streptococcus porcinus

Supervisor : Dr Jakki Cooney

Place : Laboratory of Dr Jakki Cooney ; Life Sciences Dept and MSSI ; University of Limerick ; Limerick ; Ireland

2014 Internship - 7 weeks

Title : *Microbiological control of fish aquarium filters.*

Supervisor : Yannick Ledorre

Place : UR AFPA ; Université de Lorraine ; Nancy ; France

Conferences and Schools

2022 European Summerschool in Quantum Chemistry
Università degli Studi di Palermo, Palermo, Italy

2021 Think Innovation week 2021 - Entrepreneurship and Innovation in a Global Environment
Université de Lorraine, Nancy, France

2021 Journées Théorie, Modélisation et Simulation (online)
Institut des Sciences Chimiques de Rennes, Rennes, France

2020 AEBIN Photochemistry School (online)
Donostia International Physics Center, Donostia-San Sebastian, Spain

Scientific Communication

2020 SHARPER - Note europea dei ricercatori
Università degli Studi di Palermo, Palermo, Italy

Awards and distinctions

2022 Premio Ricerc@STEBICEF 2022
Università degli Studi di Palermo, Palermo, Italy
This Call for Papers is aimed at awarding no. 5 prizes of 500.00 euro each. The Call is addressed to young PhD students, post-doctoral fellows, grant holders and contract holders who have distinguished themselves for their scientific merits by publishing, in the year 2021.

Supervision and Mentoring Activities

2022 Co-Supervisor of tow bachelor students in computer science, two months, working on the development of a VMD module facilitating the structural analysis of DNA. The *NucleicAnalysor* module is currently available on GitHub:
`https://github.com/Pitoucha/NucleicAnalysor`

Software development

2022 FindContacts

A Tcl script that adds functionalities to the VMD software. It especially allows to analyze and visualize interaction networks on a single molecule or in a complex.

https://github.com/TMiclot/FindContacts

List of published articles

(co-)First author articles

1. Tom Miclot et al. "G-Quadruplex Recognition by DARPIns through Epitope/Paratope Analogy**". In: *Chemistry A European Journal* (2022). issn: 0947-6539, 1521-3765. doi: 10.1002/chem.202201824.

2. Tom Miclot et al. "Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage". In: *Molecules* 27.10 (2022), p. 3256. issn: 1420-3049. doi: 10.3390/molecules27103256.

3. Tom Miclot et al. "Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses". In: *The Journal of Physical Chemistry Letters* 12.42 (2021), pp. 1027710283. issn: 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.1c03071.

4. Tom Miclot et al. "Forever Young: Structural Stability of Telomeric Guanine Quadruplexes in the Presence of Oxidative DNA Lesions**". In: *Chemistry A European Journal* 27.34 (June 2021), pp. 88658874. issn: 0947-6539, 1521-3765. doi: 10.1002/chem.202100993.

5. Cécilia Hognon et al. "Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses". In: *The Journal of Physical Chemistry Letters* 11.14 (2020), pp. 56615667. issn: 1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.0c01097.

Other articles

1. Emmanuelle Bignon et al. "Specific Recognition of the 5-Untranslated Region of West Nile Virus Genome by Human Innate Immune System". In: *Viruses* 14.6 (2022), p. 1282. issn: 1999-4915. doi: 10.3390/v14061282.

2. Emmanuelle Bignon et al. "Structure of the 5 untranslated region in SARS-CoV-2 genome and its specific recognition by innate immune system *via* the human oligoadenylate synthase 1". In: *Chemical Communications* 58.13 (2022), pp. 21762179. issn: 1359-7345, 1364-548X. doi: 10.1039/D1CC07006A.

3. Jeremy Morere et al. "How Fragile We Are: Influence of Stimulator of Interferon Genes (STING) Variants on Pathogen Recognition and Immune Response Efficiency". In: *Journal of Chemical Information and Modeling* (2022), acs.jcim.2c00315. issn: 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.2c00315.

4. Antonio Francés-Monerris et al. "Microscopic interactions between ivermectin and key human and viral proteins involved in SARS-CoV-2 infection". In: *Physical Chemistry Chemical Physics* 23.40 (2021), pp. 2295722971. issn: 1463-9076, 1463-9084. doi: 10.1039/D1CP02967C.

5. Antonio Francés-Monerris et al. "Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches". In: *Journal of Proteome Research* 19.11 (2020), pp. 42914315. issn:1535-3893, 1535-3907. doi: 10.1021/acs.jproteome.0c00779.

6. Cristina García-Iriepa et al. "Thermodynamics of the Interaction between the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus-2 and the Receptor of Human Angiotensin-Converting Enzyme 2. Effects of Possible Ligands". In: *The Journal of Physical Chemistry Letters* 11.21 (2020), pp. 92729281. issn:1948-7185, 1948-7185. doi: 10.1021/acs.jpclett.0c02203.

1

Introduction

The artist not only carries humanity in himself, but he reproduces its history in the creation of his work: first disorder, a general view, aspirations, dazzle, everything is mixed (barbaric era); then analysis, doubt, method, arrangement of the parts (scientific era); finally he comes back to the first synthesis, more extended in the execution.

Flaubert, Carnet 2

The genetic information determines the formation of any living being. Its integrity favors the maintenance of the individual and the conservation of the species. For this reason, genetic alteration leads on a path strewn with light and shadow. On the one hand it is the theater of evolution; the one that has led us from a single-cell life form to the dominant species on the planet. However, the alteration of genetic information can also generate harmful or dangerous effects, with pernicious consequences. There is here the mark of an apparent dualism: genetic alteration manifests itself as the catalyst stimulating the progression and transformation of living organisms, or it presents itself as the fatal herald of their decline. This last point occurs when the genetic alteration is too important, so evolution has endowed every form of life with a means of ensuring the preservation and integrity of the genetic information. If we think of the gateway to life, then the information repair system is identified with Janus, the Roman god with two faces and guardian of the city. As the gatekeeper, the first face looks outward to the environment in which individuals live, and the other face looks inward to cellular life. The repair system strives to repel the harmful effects of genetic alteration and like Janus it closes the door to the evils of life. The Janus analogy takes on a special meaning since the causes of genetic alteration originate from metabolic activity within the cells themselves, such as replication errors and the production of reactive oxygen species [1], or they originate in the external environment, of which radiation such as ultraviolet light is a good example [1]. The universe centered around the mechanism of genetic information is governed by the laws of nature, but ... « What man calls the laws of nature is nothing more than a generalization of a problem that escapes him [1]. » This is the question of the scientist, whose job is to understand and explain these phenomena. It marks the way of elucidation of the knowledge of all the richness around the genetic information, its conservation and its alteration. And this is none other than the history of the nucleic acids; biological molecule that stores the genetic information.

[1] Quote from Dohko, Golden Knight of Libra, in the Saint Seiya manga (1986).

Subject of this thesis

If the bibliography is very rich with contributions explaining the damage of canonical DNA, i.e. in the form of a double helix, it remains poorer as concerns the study of the damage on the non-canonical form of this biological structure. In this respect, this PhD thesis proposes to understand the structural influence of DNA damages on non-canonical G-quadruplex folding. Then, the work seeks to elucidate the mechanisms of interactions between proteins and damaged or non-damaged G-quadruplexes in cellular and viral systems.

CHAPTER 1

Discovery and structures of nucleic acids

Chapter contents

1.1 A brief historical reminder

The discovery of nucleic acids dates back to 1871 with the work of Johannes Friedrich Miescher, a Swiss physician and biologist, which he called « nuclein » [2]. A whole series of works followed to characterize their biological and chemical properties. Moreover the term « nuclein » was replaced under the name of nucleic acid a few years later, 1889, after the proposition of Richard Altmann. Then, the different nucleic acids were identified and their biological role became clearer [3]. In the middle of the 20^{th} century, the work of Rosalind Franklin, including the famous Photo 51, elucidated the double helix structure of DNA. It was only afterwards that James Dewey Watson and Francis Crick published a paper based on her work [4]. The identification of the Watson-Crick-Franklin interactions between nucleotides: AT and GC, followed shortly [5]. Then different interactions were identified in the work conducted by Hoogsteen [6]. The continuity of the studies on nucleic acids have allowed to highlight their biological and structural complexity. Further on, these organic molecules present an important structural polymorphism: helix [4],

triplet [7], hairpin [8], i-motif [9], and G-quadruplex [10].

1.2 Structure of nucleic acids

1.2.1 Nucleic acid bases and basepairing

Nucleic acids are biopolymers, i.e. polymers of organic molecules. More precisely, nucleic acids are a chain of nucleotides covalently linked together by a 3',5'-phosphodiester bond (Figure 1.1). Nucleotides consist of three elements: a phosphate group and a five-carbon sugar: 2'-deoxyribose (for DNA) or ribose (for RNA) and a nitrogenous base. There are five natural nitrogenous bases, which are divided into two categories: those based on a purine base (adenine, A, and guanine, G) and those based on a base (thymine, T, cytosine, C, and uracil, U). But not all of them are present for each type of nucleic acid. Thus, A, T, G and C are found in DNA. But T is not present in RNA, where it is replaced by U[1].

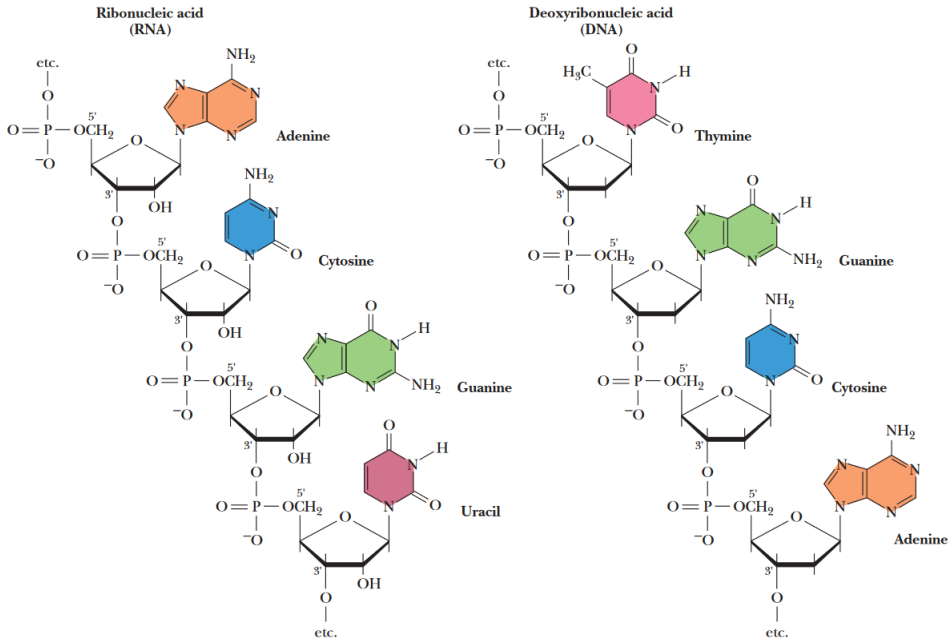


Figure 1.1: The four nucleotides that make up DNA and RNA *Reproduced from Garrett and Grisham [11].*

To fold into secondary structures, such as the double helix of DNA, nucleotides interact with each other through intermolecular hydrogen bonds (Figure 1.1). In the presence of

two strands of nucleic acid, a pyrimidine necessarily pairs with a purine base. Nucleotide make the selective pairs: G-C, which forms three hydrogen bonds, and A-T or A-U, which form only two [12, 1].

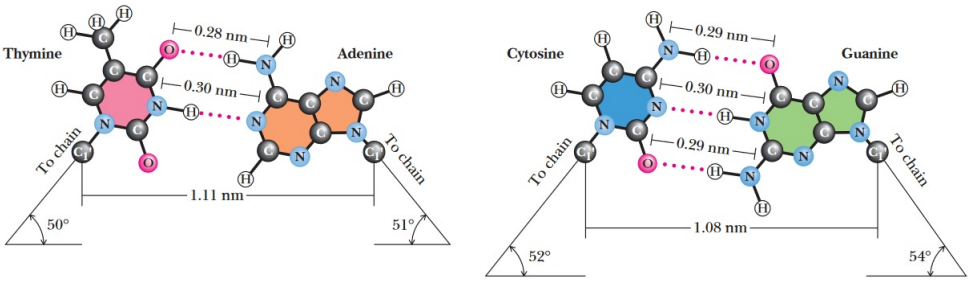


Figure 1.2: Characteristics of Watson-Crick-Franklin basepairing between nucleotides: A-T and G-C *Reproduced from Garrett and Grisham [11].*

1.2.2 Double helix: canonical structure of DNA

The best known nucleic acid structure is the DNA double helix [4]. It consists of two polynucleotide chains wrapped around each other and oriented in opposite directions. The double helix has two grooves: the major groove, which is located between the sugars of the backbones of the two chains, and the minor groove which is located between the paired bases. The presence of deprotonated phosphate groups induces a negative charge on the DNA. The base pairs globally form a hydrophobic core [13]. Also, they form π-stacking interactions that stabilize the the double helix [12].

Although seemingly simple, the double helix conformation of DNA has different variations. The most commonly encountered structure is B-DNA, but two other forms exist: A-DNA and Z-DNA. Their schematic representations are given in Figure 1.3 and their structural characteristics are given in Table 1.1.

Although the double helix structure of DNA is the most known, nucleic acids have a rich collection of secondary structures, as previously mentioned. In addition to the helical form, we may cite the triplet [7], the hairpin [8], the i-motif [9], the and G-quadruplex [10]. This last form is part of the main topic of this thesis and will be described in more detail in the following chapters.

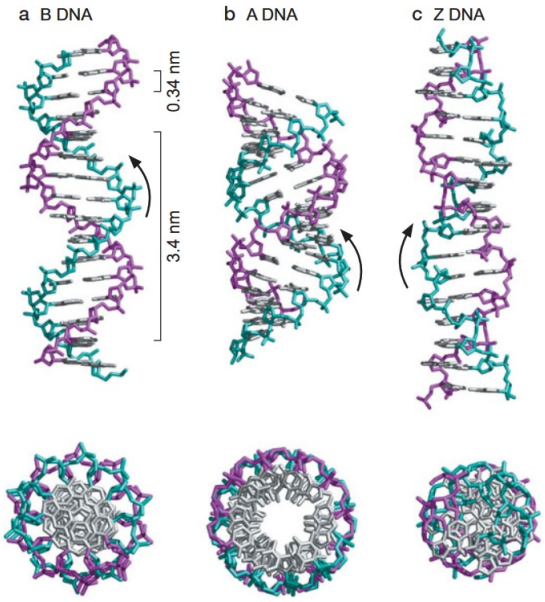


Figure 1.3: Illustration of the different conformations of the DNA double helix. *Reproduced from Watson [13].*

Structural Parameter		**A-DNA**	**B-DNA**	**Z-DNA**
Direction of helix rotation		Right handed	Right handed	Left handed
Residue per helical turn		11	10.5	12
Axial rise per residue		2.55 Å	3.4 V	3.7 Å
Pitch (length) of the helix		28.2 Å	34 Å	44.4 Å
Base pair tilt		20 °	-6 °	7 °
Rotation per residue		32.7 °	34.3 °	-30 °
Diameter of helix		23 Å	20 Å	18 Å
Configuration of glycosidic bond	dA, dT, dC	anti	anti	anti
	dG	anti	anti	syn
Sugar Pucker	dA, dT, dC	C3' endo	C2' endo	C2' endo
	dG	C3' endo	C2' endo	C3' endo

Table 1.1: Structural parameters for the different conformations of the DNA double helix. *Reproduced from Ohyama [14].*

CHAPTER 2

DNA modification: damage and mutation

Chapter contents

2.1 DNA mutation

A mutation is the modification of the genome of a living organism, it corresponds to a modification in the nucleotidic sequence of the DNA (gene mutation), or the arrangement of a chromosome (chromosomal mutation) [15], see Figure 2.1. Finally, mutations are changes that can have a positive, negative, or neutral biological impact on gene expression, transcription and phenotype [16].

2.1.1 Gene mutation

There are three types of mutations in the nucleotide sequence (point mutation) [16], namely:

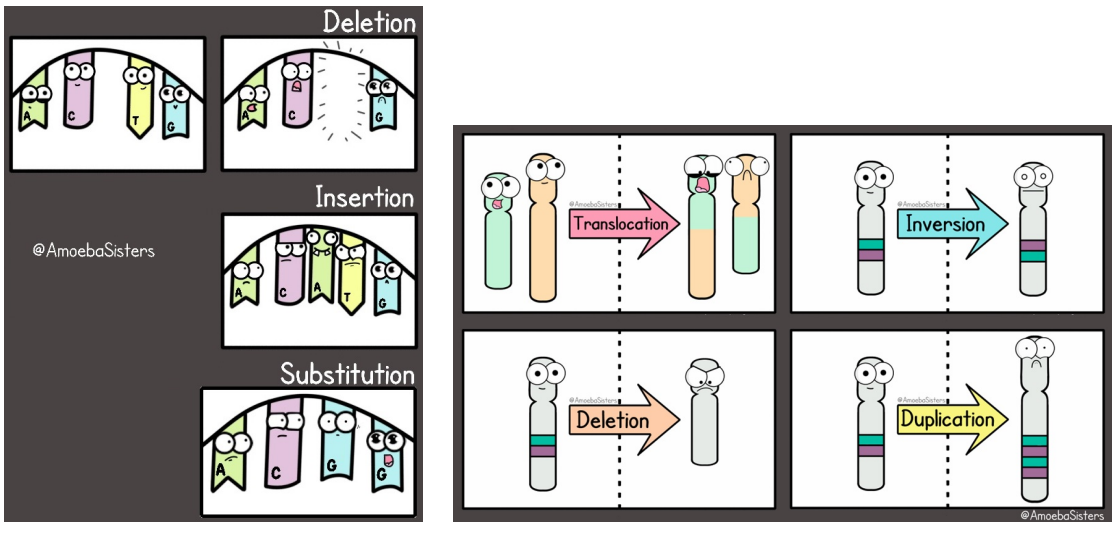
- Base deletion: Deletion of a nitrogenous base in the nucleotide sequence.

- Base insertion: Insertion of a nitrogenous base in the nucleotide sequence.
- Base substitution: One nitrogenous base is replaced by another in the nucleotide sequence. When a purine base is substituted by another purine base, or when a pyrimidine base is substituted by another pyrimidine base, an event called transition. Conversely, transversion is the substitution of a purine base by a pyrimidine base, or the substitution of a pyrimidine base by a purine base.

2.1.2 Chromosomal mutation

A chromosome mutation is a change in the arrangement of the chromosome. There are four types of rearrangements:

- Translocation: a segment of chromosome is moved on another chromosome.
- Deletion: a chromosome segment is lost.
- Inversion: the orientation of a chromosome segment is reversed.
- Duplication: a chromosome segment is doubled.



(a) (b)

Figure 2.1: Schematic representation of the different types of a) gene mutation and b) chromosome mutation. *Reproduced from the work of AmoebaSisters.*

2.2 DNA damage

DNA damage is defined as a change in the chemical structure of the nitrogenous base, the sugar or the phosphodiester bond of a nucleotide. It can be caused by the action of a mutagenic agent (chemical or physical), or be the result of a spontaneous reaction of the nucleotides. Depending on the conditions, damage may be present in a single spot, or it may form clusters of damage in the nucleotide sequence. Finally, damage is classified according to the chemical nature of the reaction product and the process that causes it.

2.2.1 Mutagenic agent induce

Base analogs It's a chemical compounds with a similarity close enough to that of the natural nitrogenous bases can be incorporated at a normal position in the nucleotide sequence. But their pairing properties are not always equivalent to those of nitrogenous bases [16]. The introduction of basic analogues can induce cell death [17, 18].

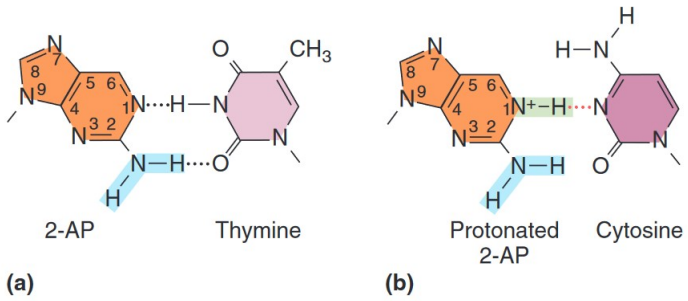


Figure 2.2: Illustration of the pairing of 2-aminopurine (2AP) with thymine and cytosine. *Reproduced from Griffiths et al. [16].*

Alkylating agents Alkylating agents are molecules that do not form permanent complexes with DNA, but modify the chemical structure of nucleotides. They add alkyl groups, sometimes ethyl or methyl, on the nucleotides. The modification causes an alteration of the pairing properties, and leads to mispairing [16, 19]. But the cells have systems that can counteract this type of damage [20, 21].

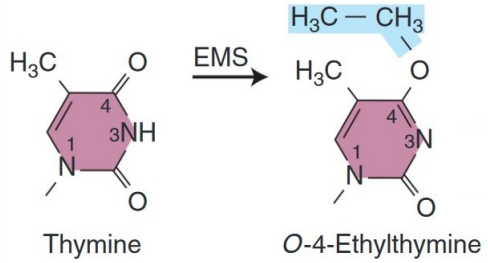


Figure 2.3: Thymine alkylation induced by ethylmethanesulfonate (EMS). *Reproduced from Griffiths et al. [16].*

Intercalating agents They are a class of planar molecules that mimic nitrogenous base pairs. They intercalate between the base pairs of the helical DNA, causing a distortion of its structure. This results in the deletion or insertion of a base pair in the next replication cycle [16, 22].

UV and ionizing radiation These are physical mutagenic agents constituted by photons of different energy. UV light mostly induces the formation of cyclobutane pyrimidine photodimer and the 6-4 photoproduct between two adjacent nucleotide bases. Here we can cite the thymine dimer whose presence correlates with the appearance of skin cancer, because these damages tend to block the replication [23, 24, 25].
Ionizing radiation causes the phosphodiester bond to break on one strand (single strand break) or on both strands (double strand breaks) of DNA [16, 26]. The product of the break in the DNA backbone gives two types of breaks. There are canonical strand break: $5' - PO_4^- / 3' - OH$, and non-canonical strand break: $5' - OH / 3' - PO_4^-$ [27, 28]. Double strand breaks are difficult to repair and hence are particularly toxic, causing high genetic instability.

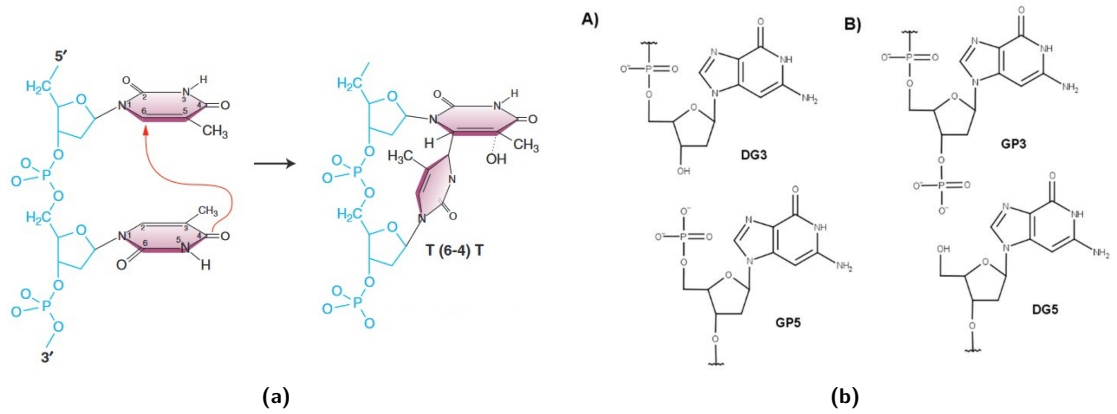


Figure 2.4: Illustration of a) the mechanism of formation of a 6-4 photoproduct, thymine dimer, induced by UV and b) products of ionizing irradiation on the backbone. *Reproduced and modified from Griffiths et al. [16] and Miclot et al. [29]*

2.2.2 DNA damage induced by the spontaneous reaction of nucleotides

Depurinatiation / Depyrimidination They correspond to the breaking of the glycosidic bond between the nitrogenous base and the sugar. If the damage occurs on a purine base (Guanine or Adenine), then it is a depurination. If the damage occurs on a pyrimidine base (thymine or cytosine), then it is a depyrimidation, the latter being

much less frequent. Both types of damage form sites where only the sugar remains in the sequence. [16, 30, 31]

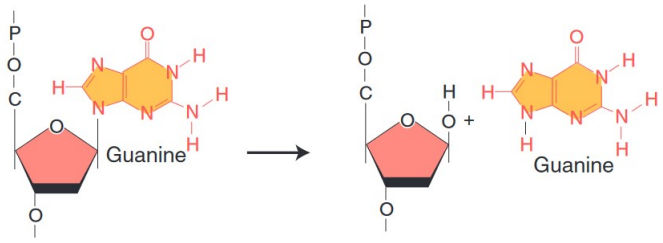


Figure 2.5: The depurination of a guanine induces the formation of an abasic site. *Reproduced from Griffiths et al. [16].*

Base deamination This is damage that results from the removal of an amine group ($-NH_3$) from a nitrogenous base. Consequently, only guanine, adenine and cytosine are sensitive to this type of reaction. For example, deamination converts a cytosine into a uracil, an RNA-specific nucleotide. [16, 32].

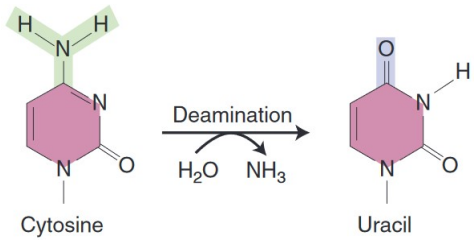


Figure 2.6: The deamination of a cytosine transforms the nucleotide into an uracil. *Reproduced from Griffiths et al. [16].*

Oxidative lesions and Reactive Oxygen Species (ROS) ROS are reactive species mainly including : superoxide radicals ($O_2^-\bullet$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($\bullet OH$). These molecules are very reactive and modify the chemical structure of nucleotides into various products [16] through oxidative reactions. Of all the nucleotides, guanine is the one with the highest reduction potential. In other words it is the most sensitive nucleotide to ROS, and gives a great diversity of reaction products which may be highly toxic [33, 34, 35].

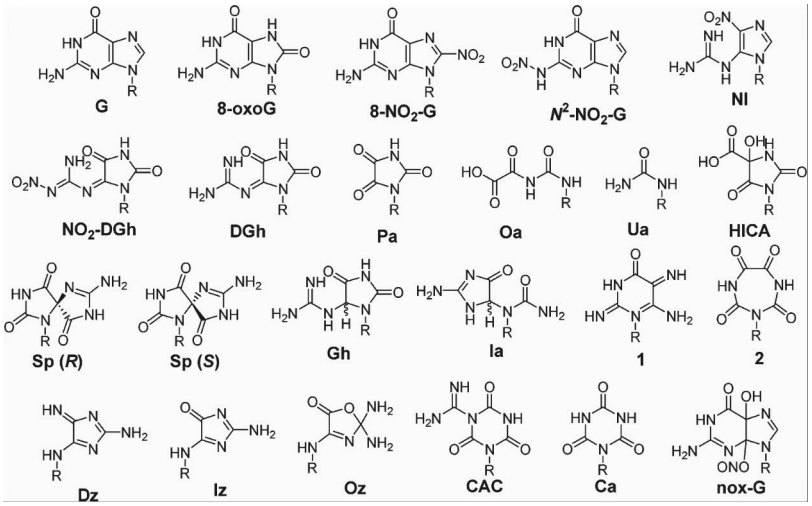


Figure 2.7: Example of some oxidation products of guanine. *Reproduced from Neeley and Essigmann [36].*

CHAPTER 3

DNA repair system

Chapter contents

To assure genetic stability, DNA is the only molecule in the cell that is not simply replaced, but is repaired when damaged. To carry out this operation cells uses several repair mechanisms [16].

3.1 Repair activity of DNA polymerase

DNA polymerases are a class of enzymes responsible for DNA replication. In addition to this function, they have the ability to identify and repair base mismatches. In reality, the replication/repair activity occurs simultaneously [37, 38, 39].

3.2 Base excision-repair (BER)

The BER repair system is one of the most important in the cell. It acts on spontaneous DNA damage: abnormal bases, apuric or apyrimidic sites (AP site), and single-strand

breaks, by involving a whole machinery of enzymes. First, DNA glycosylases, whose role is to recognize the damage and to cleave the glycosidic bond between the nitrogenous base and the sugar leading to an AP site which is recognized and removed from the nucleotide sequence by the AP-endonuclease, by cleavage of the phosphodiester bond. Then, phosphodiesterase reaches the environment of the backbone cut to remove a section of nucleotides. This "gap" is then repaired by DNA polymerase activity exploting sequence complementarity with the opposite strand [40]. BER activity is illustrated in Figure 3.1.

3.3 Nucleic excision-repair (NER)

The BER repair system is very efficient, but it is unable to intervene effectively in the presence of a large amount of damage in the nucleotide sequence or of bulky lesions. Moreover, it is unable to recognize pyrimidine dimers induced by UV light, nor damage that distorts the DNA double helix. Therefore, when one of these situations occurs, another repair system is put in place. This is the NER. In principle, it is very similar to BER, but involves a much larger protein machinery, see Figure 3.1. A deficiency of this repair system causes the appearance of two debilitating autosomal diseases: Cockayne syndrome and *Xeroderma pigmentosum* [40].

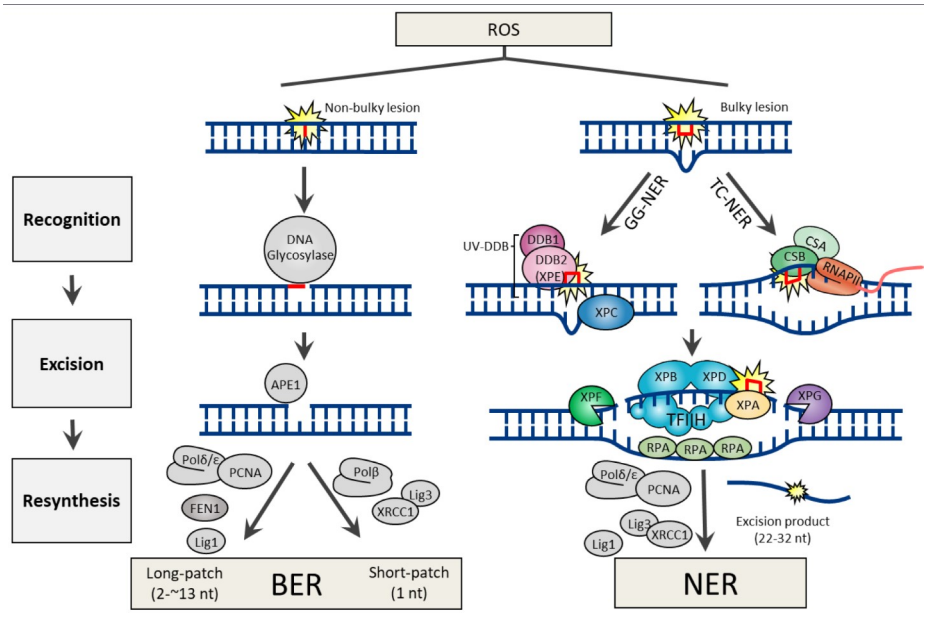


Figure 3.1: Mechanisms of the BER and NER repair systems, with their different protein contributors. *Reproduced from Lee and Kang [41].*

3.4 Double strand break repair

Double strand breaks are very toxic damage frequently leading to cell death. When this type of damage is present in the nucleotide sequence, two major repair systems can be mobilized: homologous recombination (HR) and nonhomologous DNA end joining (HNEJ). In the first case, the repair system degrades the DNA to obtain single-stranded sequences bordering the damaged area. Then it searches a healthy strand for a sequence homologous to the one present on the damaged strand. Then, the healthy complementary strand matches with the damaged strand and the activity of DNA polymerase allows the reconstruction of the nucleotide sequence. In the second repair system (HNEJ), proteins will degrade the DNA to obtain 5'-P and 3'-OH ends, which are recognized and repaired by DNA ligase IV [42, 43].

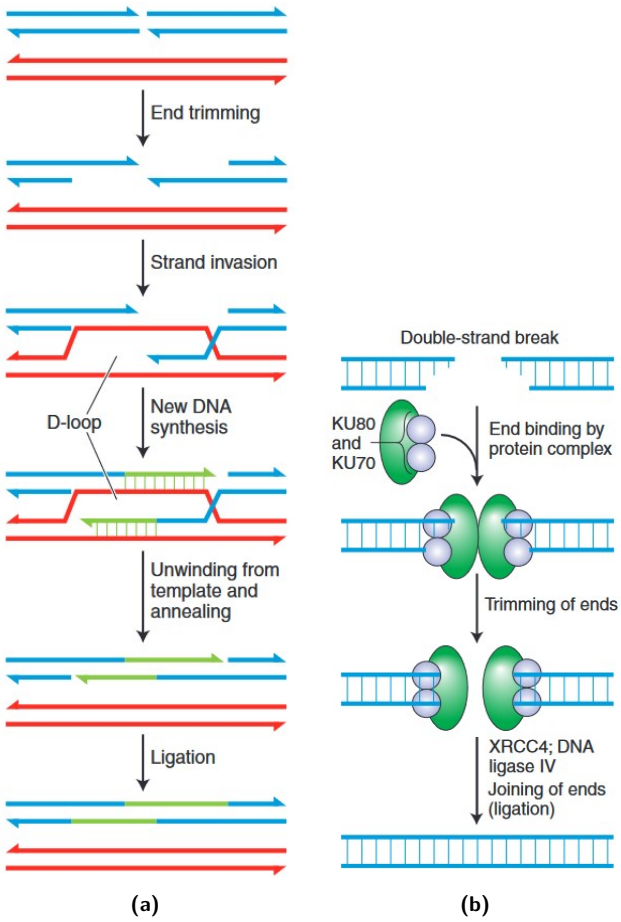


Figure 3.2: Mechanism of double strand break repair a) by homologous recombination and b) by the HNEJ system. *Reproduced from Griffiths et al. [16] and Miclot et al. [29]*

CHAPTER 4

G-quadruplex: non-canonical structure of nucleic acids

Chapter contents

4.1 G-quadruplex and bibliometric impact

Among non-canonic arrangements of nucleic acids G-quadruplex shows a growing interest. Indeed, an analysis of the literature in the *Web of Science* database shows an almost exponential increase in the proportion of publications presenting the keyword « G-quadruplex » , see Figure 4.1; the same is true for the number of citations to articles on this subject. More precisely, G-quadruplex research area produced few publications between the years 1990 and 2000. But, in the period between 2001 and 2006, the number of publications and citations has been steadily increasing. Consequently, starting in 2007 and continuing until now, G-quadruplex represent a research hot topic with a steep growth.

The importance of G-quadruplexes is not only revealed by the increase in the publications that mention them. In fact, a complex network is articulated around this research topic. For example, it is possible to shed light on the universe established around the research on G-quadruplexes, using the VOSviewer software [44]. This software creates bibliometric networks from the analysis of search results conducted in a database. The

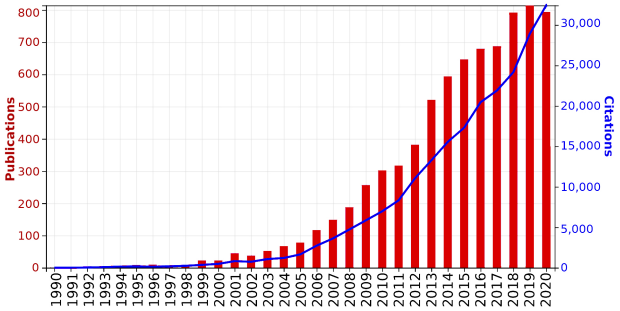


Figure 4.1: The analysis of the bibliography by the *Web of Science* database returns a growing trend in the number of publications and citations involving the keyword « G-quadruplex ». Citation Report graphic is derived from Clarivate *Web of Science*, Copyright Clarivate 2021. All rights reserved.

result of the analysis of the abstract of the articles from the previous research on the *Web of Science* database is represented in Figure 4.2. There is a visual representation of an important network involving G-quadruplex-related thermes. All of them are grouped into four distinct clusters: cell biology, genetics, G-quadruplex structure, and detection and biotechnology applications. Thus, the scientific area interested in the non-canonical G-quadruplex form of nucleic acids is growing and showing a large area of wealth.

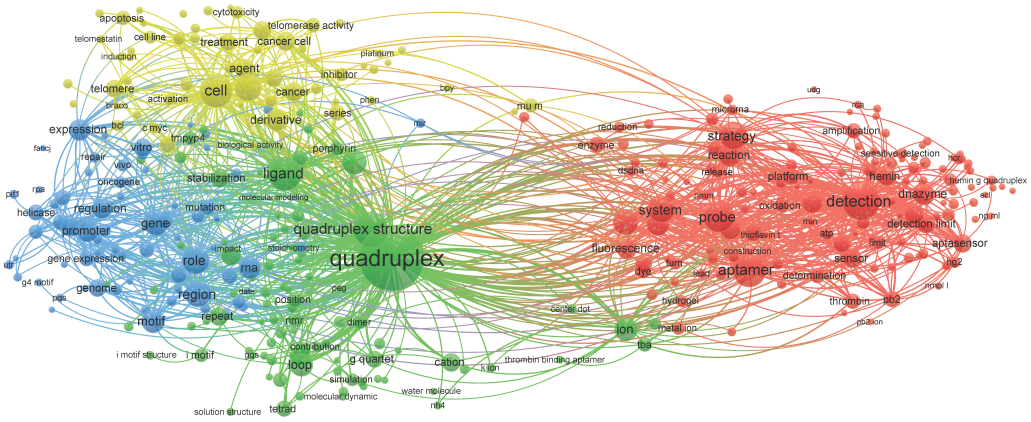


Figure 4.2: The universe built around the G-quadruplex is very rich and composed by four large clusters, represented here by different colors. Yellow : cell biology, Blue : genetic, Green : structural biology, Red : detection & biotechnologies.

4.2 Biological function of G-quadruplexes

To understand the biological importance of G-quadruplexes, it is necessary to recall that not all the nucleotide sequences are involved in the replication and transduction of the genetic information. Indeed, some sequences are involved in the regulation of genetic mechanisms [1]. The regulatory sequences can be presented in a canonical form, the so called double-helical or B-DNA, or in non-canonical forms such as G-quadruplexes. The biological role of the latter is, however,very subtle both considering their biological role

and the interactions with other non-canonical structures. On this point, we can mention the interdependence of the stability of G-quadruplexes and i-motifs: the stabilization of G-quadruplexes leads to the destabilization of i-motifs, conversely the stabilization of i-motifs leads to the destabilization of G-quadurplexes [45].

G-quadruplexes are involved in various cellular processes, such as transcription, recombination, replication and regulation of the cell life cycle [46, 47]. Telomeric G-quadruplexes (h-Telo) are found at the ends of chromosomes, in the telomeres, and their formation is controlled by telomere end-binding proteins [48]. h-Telo G-quadruplex are involved in the regulation of telomerase activity. Overexpression of telomerase is often linked to the immortality of malignant cells. Therefore, the understanding of the structure and dynamics of G-quadruplexes is fundamental to develop appropriate therapeutic strategies in cancer research [49]. From a more general medical point of view, the study of G-quadruplexes is also of interest, for exemple Asamitsu et al. [50] showed their role in the modification of nerves. Furthermore, they are responsible for the development of neurological diseases, including the X-linked intellectual disability syndrome [51].

In a virological context, G-quadruplexes also represent an interesting therapeutic target because they play a role in the viral replication cycle [52, 53, 54]. For example, the SARS unique domain (SUD) of the non-structural protein of SARS-CoV, and SARS-CoV2, can bind to guanine-rich RNA sequences, including G-quadruplex [55, 56], which allows the virus to evade the host immune system's response [57]. In the same context, G-quadruplex RNA structures were identified in the SARS-CoV-2 genome [58]. G-quadruplexes are present in the diversity of life. They are part of the resistance mechanism of the bacterium *Deinococcus radiodurans* towards ionizing radiations [59] and are involved in the mechanisms of plant development and growth [60, 61]. So, the interest of G-quadruplexes is so wide that bioinformatic algorithms and tools have been developed for the sole purpose of tentatively identifying and locating them in the genome [62, 63].

4.3 Biotechnological interests of G-quadruplex

G-quadruplexes represent an interesting topic in the development of nanotechnologies. Some G-quadruplexes are lipophilic, so they can be inserted into membranes and mimic ion channels [64]. They are also able to recruit hemes into cells, moreover the heme/G-quadruplex association forms a DNAzyme with peroxidase activity [65, 66]. They can also be used in the assembly of nanotubes [67]. Other biotechnological applications involve ionophores, hydrogel, enhanced fluorescence, assisted polymerization of crystalline nanorods, DNA Origamis or DNA logic gates [68, 69].

2

Structural parameters of G-quadruplexes

CHAPTER 5

Structures of G-quadruplexes

Chapter contents

5.1 Properties of the guanine and formation of G-quadruplexes

Guanine is an aromatic organic molecule with the chemical formula: $C_5H_5N_5O$. It is one of the nitrogenous bases found in nucleic acids (DNA or RNA). This molecule adopts a planar structure, which is derived from purine. Figure 5.1 shows the chemical structure of guanine. The proper nomenclature is as follows [70]:

- when the base only is referred then the name guanine should be used, because the R moiety is a hydrogen (H);

- when discussing the nucleoside then the name guanosine, or deoxy-guanosine, should be used, because the R moiety is a sugar (ribose or deoxyribose);
- when referring to the nucleotide then name guanylic acid, or deoxy-guanylic acid should be used, because the R moiety is now a sugar (ribose or deoxyribose, for RNA or DNA, respectively) covalently linked to a phosphate.

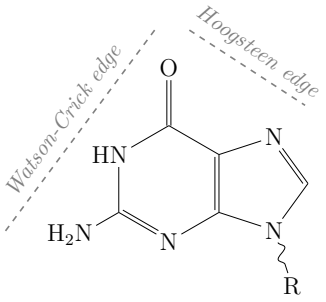
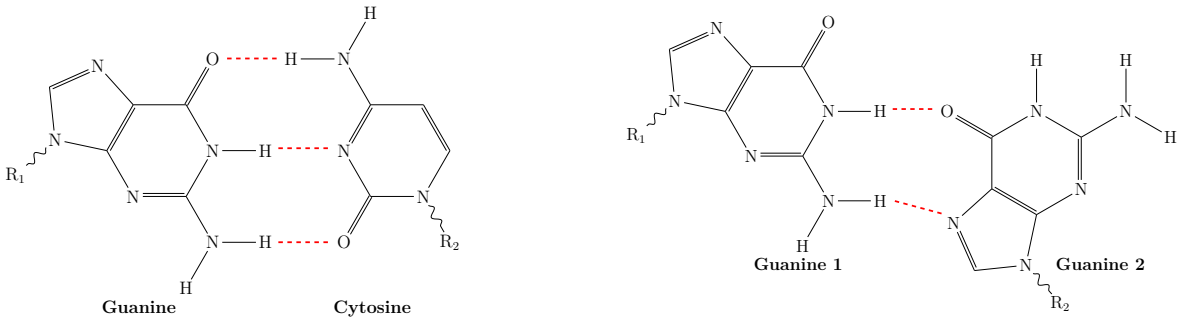


Figure 5.1: Chemical structure of the guanine, with the representation of the Watson-Crick edge and Hoogsteen edge involve in base pairing. *Adapted from Leontis [71]*

Guanine present two most prominent patterns leading to the formation of hydrogen bonds with other nucleotides. The first was described by Watson and Crick, while the other was described later by Hoogsteen [4, 6]. Each of the two interaction sites is shown in the Figure 5.1. Due to their different location, these two sites generate hydrogen matching of different organisation. The Watson-Crick area can make three H-bonds, and the Hoogsteen area can only make two H-bonds. Consequently, this arrangement geometrically limits the possibilities of matching with other nucleotides. The Watson-Crick matching of guanine is complementary to a cytosine, while the Hoogsteen matching changes this rule and grants guanine the ability to bind other nucleotides. Donohue and Trueblood [72] showed the possibility of an adenine-guanine pairing, which was then taken up and discussed further by Hoogsteen [6]. Nevertheless, a guanine is also able to form hydrogen bonds with another guanine, as shown in Figure 5.2. Although Watson-Crick type base pairing is the most commonly encountered, the Hoogsteen model also exists in double-stranded DNA, albeit in smaller proportions. Interestingly, Hoogsteen base pairs also have a biological role in damage and repair, in DNA/protein interactions and even in the replication machinery [73, 74].

Watson-Crick and Hoogsteen base pairs represent two different types of matching. They can also act in concert in the case of triple helix structure of nucleic acids [75]. Reverse Watson-Crick and Hoogsteen base pairing act in the formation of i-motifs, cytosine-rich four-stranded nucleic structures [76]. However, there are particular structures that



(a) Watson-Crick base pairing

(b) Hoogsteen base pairing

Figure 5.2: Hydrogen bonds of guanine with other nucleotides, according to the Watson-Crick or Hoogsteen models. H-bonds are shown as dotted red lines.

are formed by Hoogsteen type pairing only. It involves four guanines, in the same G-G configuration as shown in Figure 5.2, which establish an almost planar arrangement. This particular configuration was already observed *in vitro* in 1962 [77], and is called a guanine tetrad, or quartet. But these quartets are not found alone in a G-quaduplex. They are stacked on other quartets $\pi - \pi$ stacking interactions, forming a system of several parallel quartets. A guanine quartet arrangement and a schematic representation of a three-quadrtets G-quadruplex is shown in Figure 5.3.

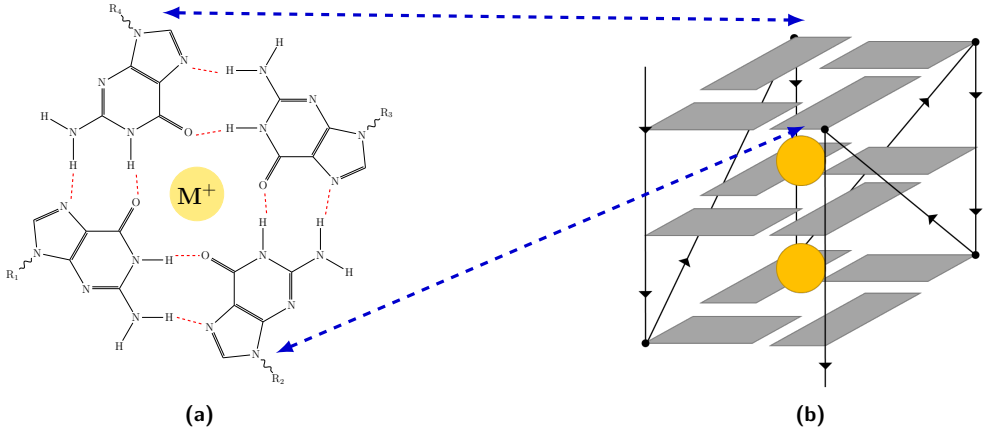
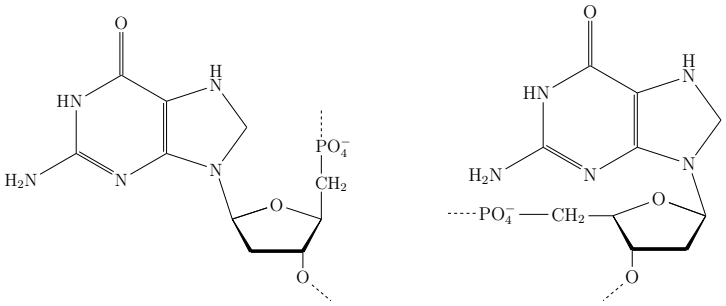


Figure 5.3: Guanine quartet **(a)** superimposed on two other quartets **(b)** in a G-quadruplex arrangement stabilised by the presence of alkali metal cations (M^+ represented by yellow spheres).

Hoogsteen pairings in guanine quartets can be formed in *anti* or *syn* arrangement with respect to the glycosidic bond (Figure 5.4) [78]. If a quartet undergoes alteration of the hydrogen bond pattern, such as a decrease of H-bonds, then the destabilisation of the quartet usually follows [79].

However, the stabilisation of a G-quadruplex structure does not rely solely on Hoogsteen base pairs. In fact, a G-quadruplex consists of at least two guanine quartets parallel to each other and including an alkali metal cation. The presence of the cation is necessary



(a) *Anti* **(b)** *Syn*

Figure 5.4: *Anti* and *Syn* conformations of a deoxy-guanylic acid fragment.

to counterbalance the negative charge of the phosphate groups in the sugar-phosphate backbone. The most common cation is potassium, K^+, but there are other types of cations within G-quadruplexes, such as alkali and alkali earths metals and sometimes non-metals, such as the ammonium ion, $NH_4{}^+$. Despite this rich variety, not all cations have the same stabilising strength [80, 81]. Largy, Mergny, and Gabelica [82] propose the following order: $Sr^{2+} > K^+ > Ca^{2+} > NH_4{}^+, Na^+, Rb^+ > Mg^{2+} > Li^+ \geq Cs^+$. However, the presence of a cation is crucial for the formation and stabilization of G-quadruplexes [78].

Two last important parameters for G-quadruplex stability and formation are: hydration and glycosidic bonds angles. Indeed, the hydration state controls the conformation, also called topology, during the folding of a nucleic acid into a G-quadruplex. More precisely, it is the dehydration that allows this specific folding. Depending on nucleotide sequence, length of loops [83], and environmental conditions, a G-quadruplex will fold in one of three possible topologies: parallel, antiparallel and hybrid (also call mixed parallel/antiparallel, or 3+1) [84, 85, 86]. Water molecules interact mostly with N2, N3 and O4' atoms of guanine [78]. Also, we know that the glycosidic bonds angles differ in the G-quadruplexes topologies [87]. Therefore, the environmental conditions capable of modifying these angles will also modify the global G-quadruplex topology.

5.2 Topology classification

G-quadruplexes are cation-stabilised stacks of guanine quartets, connected by sugar-phosphate loops. However, different DNA sequences produce different arrangements, i.e. conformations in which each nucleotide adopts different geometrical parameters. Factors

affecting the topology include the number of molecules forming the G-quadruplex, the organisation of the loops, and the arrangement of the quartets. Some experimental techniques, such as circular dichroism, are able to provide some information on the topology of G-quadruplexes. However, resolving at an atomistic scale the precise structures of G-quadruplexes is more delicate and time-consuming.

5.2.1 Bulge formation between quartets

The sequence of G-quadruplexes often consists of repeating motifs: for example `TTAGGG` for the human telomeric G-quadruplex. For this reason, G-quadruplex sequences are often represented with the guanines forming the quartets separated by linking sequences, such as in the consensus sequence: $G_{3+}N_{L1}G_{3+}N_{L2}G_{3+}N_{L3}G_{3+}$; where N_L is a nucleic sequence of arbitrary length [88, 89]. However, it has been seen that G-quadruplexes can be formed even if there are breaks in the sequence of three guanines, for example in `d(GGG TT GCGG A GGG T GGG CCT)`, where a cytosine interrupts the guanine sequence. The latter is however ejected from the G-quadruplex core, forming a small loop called a bulge (see Figure 5.5) [90, 89, 91].

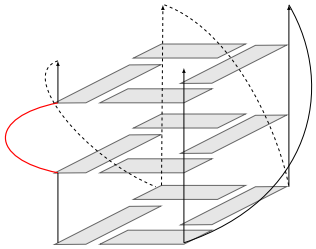
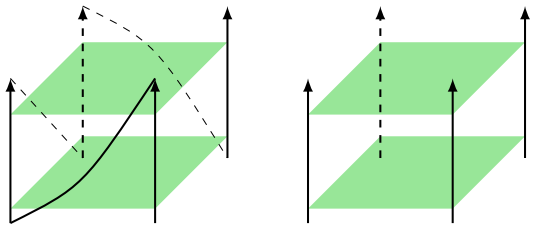


Figure 5.5: Intramolecular G-quadruplex containing a bulge, in red.

5.2.2 *Intra-* and *inter-* molecular folding of G-quadruplexes

G-quadruplexes form a four-stranded DNA or RNA superhelix, and thus four grooves. However, not all G-quadruplexes have the same topology. G-quadruplexes can be formed by the folding of a single DNA strand, in which case they are said to be intramolecular (or unimolecular) or of different DNA strands, in which case they are said intermolecular. For example bimolecular (dimeric) or tetramolecular (tetrameric), are formed by two or four different DNA strands, respectively [88], as shown in Figure 5.6.

Each G-quartet in a G-quadruplex can be constituted, theoretically, by 1 to 4 DNA strands, which we will name sequentially from A to D. In the example of a dimeric G-quadruplex: either two guanine bases in the tetrad belong to strand A and the other two



(a) Intra-molecular **(b)** Inter-molecular

Figure 5.6: Schematic representation of *intra-* and *inter*-molecular G-quadruplexes. Each green tetragon represents a quartet.

to strand B (A2-B2), or three guanine bases belong to strand A and only one to strand B (A3-B1). From this reasoning, it is possible to summarize in Table 5.1 all possibilities to form intramolecular and intermolecular G-quadruplexes.

Topology	A	B	C	D
Monomeric	4			
Dimeric	3	1		
	2	2		
Trimeric	2	1	1	
Tetrameric	1	1	1	1

Table 5.1: Possible foldind of a G-quadruplex with a maximum of four strands, from A to D. Each number represents the number of guanine bases in a G-tetrad belonging to a strand in the G-quadruplex.

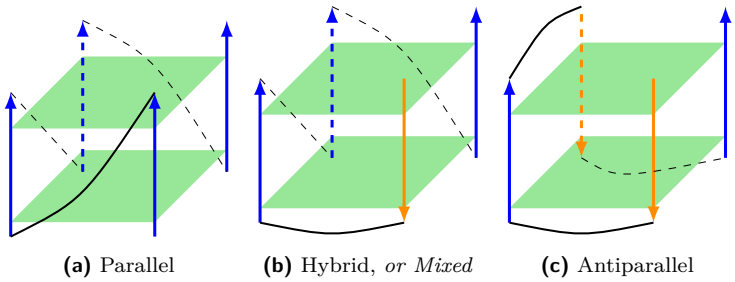
5.2.3 G-quadruplexes topologies definitions

Previously it was mentioned that three topologies exist for a G-quadruplex: parallel, hybrid (or mixed), and antiparallel. These nomenclature originate from the relative arrangement of the ($5' \rightarrow 3'$) directions of the four strands connecting the stacked G-tetrads [78], see Figure 5.7.

- If all strands forming the guanine tetrads are oriented following the same polarity with respect to each other, then the G-quadruplex is parallel. (↑↑↑↑)
- If only one strand has a reversed polarity orientation with respect to the others, then the G-quadruplex is hybrid. (↑↑↑↓)
- If two strands are oriented with reversed polarity with respect to the other two, then the G-quadruplex is antiparallel. (↑↑↓↓)

5.2.4 Linking loops types

In intramolecular G-quadruplexes the guanine bases in the G-tetrads are linked by loops made by a sequence of nucleotides that do not participate in the formation of

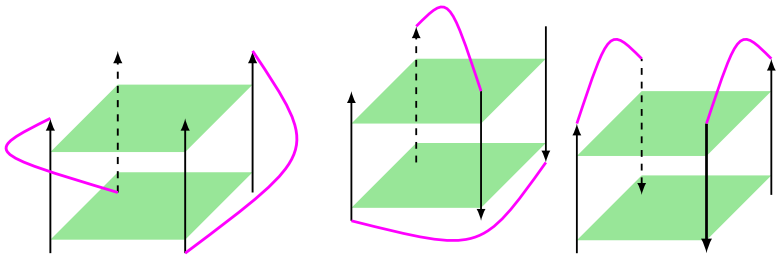


(a) Parallel **(b)** Hybrid, *or Mixed* **(c)** Antiparallel

Figure 5.7: Schematic representation of the three possible topologies in an intramolecular G-quadruplexes, where the arrows represent the ($5' \rightarrow 3'$) strand direction. Arrows with the same color represent parallel strands with the same direction.

the same guanidic core of the G-tetrad. There are three possible types of loops (see Figure 5.8): external, diagonal and lateral [78].

- External loops are established between two adjacent parallel strands.
- Diagonal loops connect two opposite anti-parallel strands.
- Lateral loops connect adjacent anti-parallel strands.

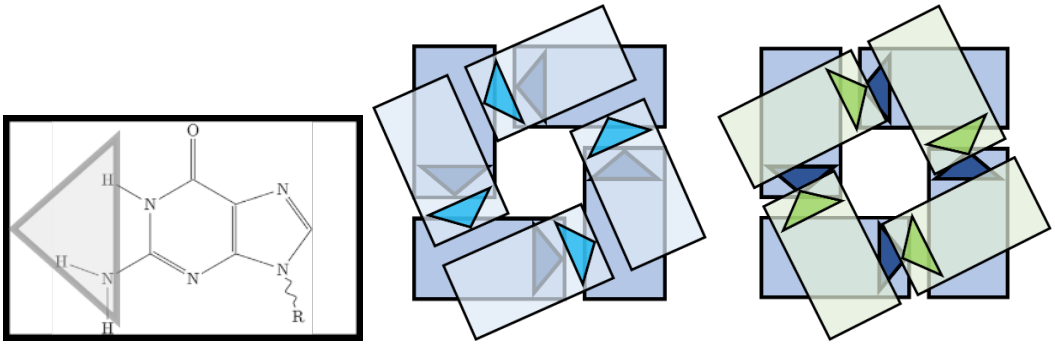


(a) External, or *Propeller* **(b)** Diagonal **(c)** Lateral

Figure 5.8: Schematic representation of the three loop arrangements (in magenta) in G-quadruplex. This examples are shown in inter-molecular G-quadruplexes.

5.2.5 Quartets stacking and classification

The parallel superposition of quartets is the core motif leading to the G-quadruplex formation, originating by a $\pi - \pi$ stacking interactions between the guanine bases. In each quartet, the guanines are arranged according to a direction of rotation, sometimes called quartet polarity, as we discussed in the previous section and in Figure 5.3: either the two superimposed quartets show homopolar stacking, i.e. the four guanines in the two tetrads rotate in the same direction; or they show heteropolar stacking, i.e. the guanines of the two tetrads rotate in opposite directions [85, 92], see figure Figure 5.9.



(a) Schematic guanine polarity **(b)** Homopolar stacking **(c)** Heteropolar stacking

Figure 5.9: Schematic representation of the stacking of two G-tetrads. The O6 oxygen is always near the center of the tetrad. *Adapted from Lech, Heddi, and Phan [92].*

CHAPTER 6

Structural Elucidation by Circular Dichroism

These days, every Tom, Dick and Harry thinks he knows what a photon is, but he is wrong.

Albert Einstein – 1951

Chapter contents

The use of circular dichroism (CD) properties of polarized light gives rise to a very powerful molecular analysis technique. By exploiting the optical activity of chiral molecules, or supramolecular aggregates such as proteins or nucleic acids, it is possible to characterize their secondary structures and monitor their conformational changes, for example following their interactions with ligands. This chapter proposes to recall some fundamental aspects of CD.

6.1 Wave properties of light

In the history of physics, light has been a subject that has attracted many intellectuals from all over the world. After the ancient and medieval theories, there was a debate lasting for some centuries opposing the corpuscular theory enunciated by Newton and the wave theory enunciated by Huygens [93]. Today, we know that light has both properties, i.e. it is both a wave and a particle.

The photon is the elementary particle of light, however, light can be described as an electromagnetic field. For more clarity, a field is what represents all the quantitative values (q) that a physical entity can take, at any point in a given space (x,y,z) and time (t) [94, 95]. There is a spatio-temporal dependence (x,y,z,t) of the constituent values of a field. Thus a field is a five-dimensional object, which is constituted by the set of all (x,y,z,t,q) possible coordinates.

More specifically, an electromagnetic field is the sum of an electric field ($\overrightarrow{E}$) and a magnetic field ($\overrightarrow{M}$), perpendicularly to each other. Both propagate perpendicularly to the axis of their direction of propagation ($\overrightarrow{d}$), as represented in Figure 6.1. In fact, this direction of propagation is none other than that of light. Electric and magnetic fields have a wave-like behaviour. As a result, an electromagnetic field behaves like a wave [96]. The beginnings of the formulation of the properties of electromagnetism were formalized by James Clerk Maxwell in the 19^{th} century [97]. Today the wave equations describing electromagnetic properties are presented in four differential equations [94, 98].

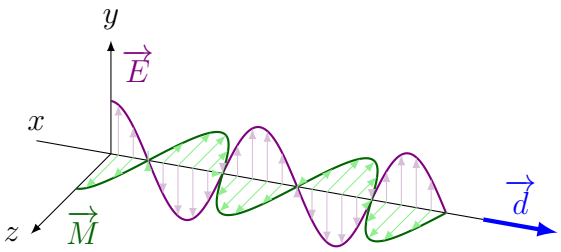


Figure 6.1: Perpendicular propagation in the direction $\overrightarrow{d}$ of an electromagnetic wave, composed of a magnetic wave $\overrightarrow{M}$ (in green) and an electric wave $\overrightarrow{E}$ (in purple), perpendicular to each other.

Here, we will focus only on the following properties of light waves:

- Light is an electromagnetic wave.
- Light can be polarized.

- The speed of light in vacuum c is exactly 299 792 458 $m.s^{-1}$ [99].

- The wavelength λ, which is the distance between two point (usually taken as the wave maximum amplitude) of the oscillation in phase. The energy of a photon is directly related to the wavelength of light.

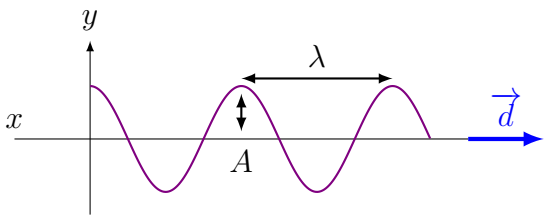


Figure 6.2: A wave is characterized by its wavelength λ and its amplitude A.

- The frequency f, which is dependent on the speed of the light in vacuum c and the wavelength of the light λ :

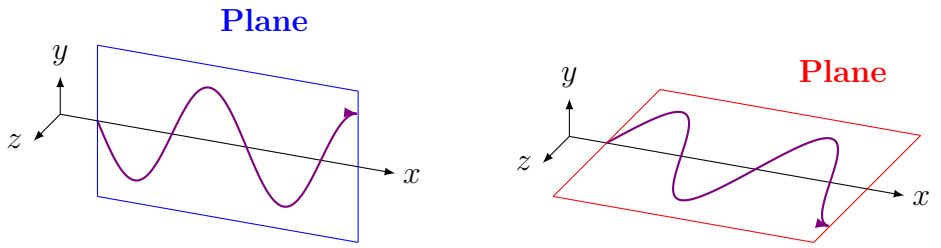
$$f = \frac{c}{\lambda} \tag{6.1}$$

6.2 Light polarisation

Light is a wave that it can be polarized, i.e. it can be oriented in space. We can explain this phenomenon if we consider that the oscillation of the electric wave in the direction d (here, we consider that the direction corresponds to the x-axis) is the result of two components able to oscillate separately from each other and at a given frequency. We called this two components : y component and z component. Finally, we notice that, this reasoning dismantle the simple vectorial picture of the light wave [100].

6.2.1 Linear polarisation

When the oscillations of the y and z components reach a maximum at the same time they are said to be "in phase". So when the vibrations of the y and z components are in phase, or in phase opposition, and have a certain proportionality to each other, then the result is a direction in the polarisation plane P_{xz} [100, 98]. There are several polarization planes generally ranging from $+90°$ to $-90°$. A $0°$ plane is a vertically polarized light and is perpendicular to a plane at $+90°$ or at $-90°$, corresponding to a horizontally polarized light. So there is an equivalence of orientation only between the $+90°$ and $-90°$ planes.

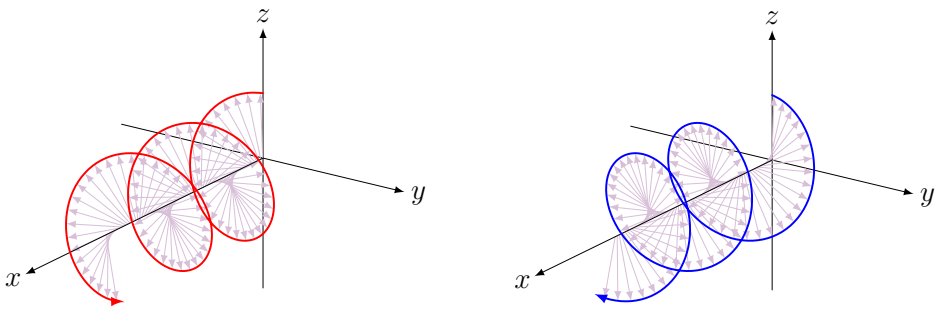


(a) Linear polarisation oriented at 0° **(b)** Linear polarisation oriented at −90°.

Figure 6.3: 0° polarization plane resulting to $y_{vibrations} = 1$ and $z_{vibrations} = 0$, and +90° polarisation plane resulting to $y_{vibrations} = 0$ and $z_{vibrations} = 1$ [100]. These two planes are perpendicular to each other.

6.2.2 Elliptical polarization and circular dichroism

The polarization of light does not always follow a plane, it can be elliptical if the oscillations of the y and z components are not in phase and present different amplitudes, leading to elliptically polarized light. But sometimes the phase shift of the y and z component vibrations is approximately 90° and happens with the same amplitude. In this case the polarization becomes circular, with two directions of polarization. If the electric wave describes a circle rotating clockwise with an angular velocity α, then it's named right-hand circular polarized light. Conversely, if the electric wave draws a counter-clockwise circle we are in presence of left-hand circular polarization [96, 98].



(a) Left-hand circular polarization **(b)** Right-hand circular polarization

Figure 6.4: Right-handed and left-handed circular polarization of light.

To conclude, the designation « Circular dichroism (CD) » refers to the circular polarisation of the light wave. Since there are only two possible oscillations, left or right, the designation is also based on the term « dichroism » which comes from the Greek word *dikhroos* which means « of two colors » [101]. A molecule is said « dichroic » when it absorbs left and right polarized light differently [102]. Finally, according to Fresnel's construction, we can consider that the resultant of two circularly polarized waves is a linearly polarized wave.[96]. There is an intrinsic link between linear and circular polar-

ization, since the sum of two left- and right-handed polarized lights, in phase and with the same amplitude, produces a linear polarized light.

6.2.3 « Natural » light

Light is typically non-polarized, which means that all overlapping waves sent by the source are polarized according to different and random planes [96].

6.3 Optical activity and chirality

The optical activity is the ability of a substance to change the plane of polarization of the incident polarized light. This property can be rationalized by the symmetry of the molecule (or macromolecule), indeed it is related to the absence of reflection symmetry [100]. Maurizot [96] goes into more details and lists the three geometrical conditions that a molecule must possess in order to be optically active. The molecule in question must have neither center of symmetry, nor plane of symmetry, nor improper axis of symmetry. When we pay attention to these details, we realize that these are the necessary conditions for a molecule to be chiral. As a reminder, two molecules with the same raw formula, but a different topology are called isomers. When two isomers are the mirror image of each other, and their structure is not superimposable, then they are chiral. The two mirror images of the molecule form a pair and are called enentiomers [103]. For example, the amino acid alanine, in Figure 6.5, is a chiral organic molecule.

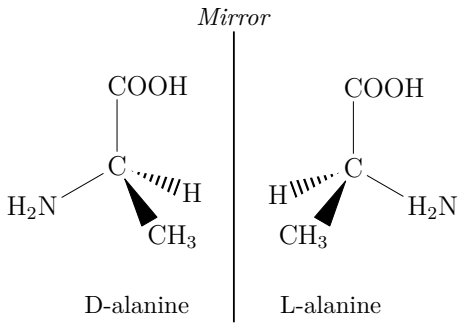


Figure 6.5: The amino acid alanine is a chiral organic molecule. It has two configurational isomers: D-alanine and L-alanine. One is the mirror image of the other, and are not superimposable. So D-alanine and L-alanine are enantiomers, with the internal C as chiral center.

Optical activity takes place only for one enantiomer of a chiral molecules. To explain this phenomenon, we must remember that linearly polarized light is a resultant of left and right circularly polarized light waves (respectively $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$). If the optical path in

a solution is different for $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$. Then the solution has a circular birefringence, i.e. different refractive indices for right and left circularly polarized light. Then a phase shift between $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$ may be observed [96]. Concretely, the incidence of the polarized light on the molecule according to the angle α induces a down-up movement of the electrons in the direction α. Then, it will generate an electric field which comes out of the molecule. In other words, if the electric wave of the light enters in contact with the molecule at the point P, it will come out at the point $P + D$, with D the diameter of the molecule [100]. Consequently the phase shift between $\overrightarrow{W_l}$ and $\overrightarrow{W_r}$ depends on the diameter of the molecule and by its capacity to absorb preferentially $\overrightarrow{W_l}$ or $\overrightarrow{W_r}$. Now, by applying the Fresnel construction to $\overrightarrow{W_l}$ or $\overrightarrow{W_r}$ dephased, we have a polarized light in another plane [96]: $\alpha \pm \beta$, as shown in Figure 6.6 below.

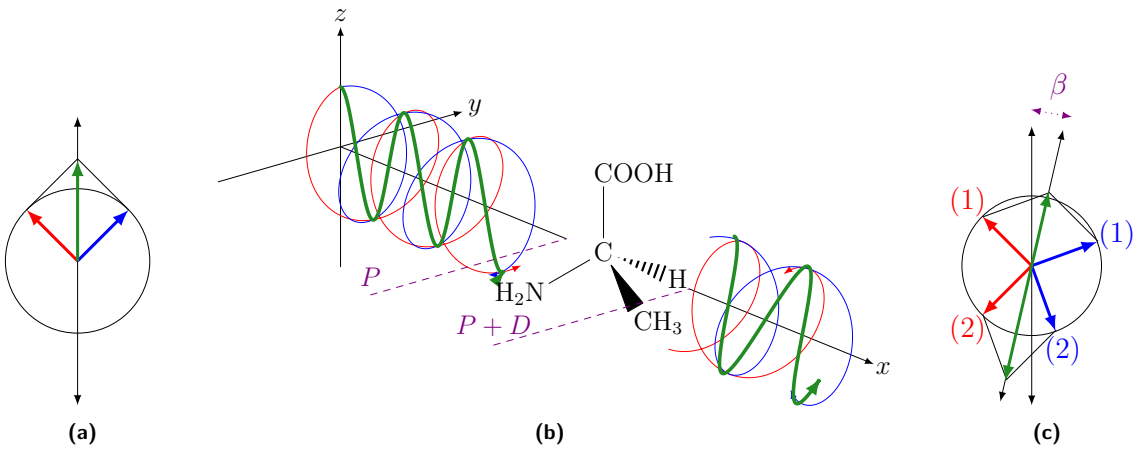


Figure 6.6: **(a)** Fresnel construction showing the superposition of $\overrightarrow{W_l}$ (red) and $\overrightarrow{W_r}$ (blue) forming a linear polarized light according to the angle *alpha* (here, equal to +90°). **(b)** This light beam is send to a D-alanine molecule. The ray $\overrightarrow{W_r}$ is absorbed by the molecule at point P and comes out at point $P+D$, while $\overrightarrow{W_l}$ is not absorbed. So $\overrightarrow{W_r}$ has a delay and is out of phase with $\overrightarrow{W_l}$. **(c)** This causes a shift of the polarization plane to the right, as shown by a second Fresnel construction. *Based on Maurizot [96] and Feynman, Sands, and Leighton [100]*

6.4 Circular dichroism

The phenomenon of circular dichroism is intimately linked to absorption and optical activity. Since it is only observed in the spectral range absorbed by an optically active molecule. But unlike absorbance which is always positive, circular dichroism can have positive and/or negative values. In fact, the circular dichroism is explained by developing what has been mentioned in section 6.3. An optically active molecule preferentially

absorbs left or right polarized light. This generates a phase shift of the two components, thus a polarization of the light according to a different angle plane than that of the incident light. However, if we go back to the Figure 6.6 we notice that the discussion does not mention the intensity of the absorption of light by the molecule. Indeed, we see in c) that the Fresnel diagram uses vectors of the same radius. But a molecule will not always send back an absorbed light with the same amplitude. That is to say that there is an intensity of absorption. When $\overrightarrow{W_r}$ and $\overrightarrow{W_l}$ are projected with different radii [96], the polarization is not linear, but elliptical as we can as can be seen in Figure 6.7 below.

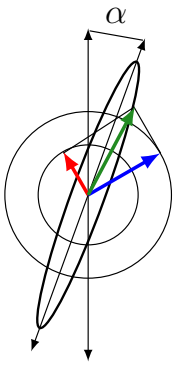
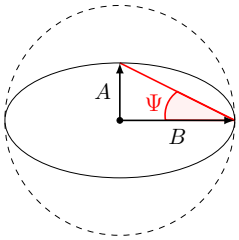


Figure 6.7: The construction of a Fresnel diagram allows to explain the elliptical polarization of a light resulting from an absorption intensity of $\overrightarrow{W_r}$ (blue) and $\overrightarrow{W_l}$ (red) by a molecule. The alpha angle is the rotation of the major axis of the ellipsis with respect to the polarisation axis of the incident light [96].

Two specific parameters are used to describe the circular dichroism phenomenon of a molecule. The first is the *dichroic molar absorption coefficient* $\Delta\varepsilon$, which is expressed in $mol^{-1}.cm^{-1}$. $\Delta\varepsilon$ is calculated as the difference between the absoption coefficients of left and right polarised light : respectively ε_l and ε_r. So its formula is :

$$\Delta\varepsilon = \varepsilon_l - \varepsilon_r \tag{6.2}$$

The last parameter is called ellipticity. In order to explain what this is, it is worth recalling that the light incident on the sample is circularly polarised, while the light exiting the sample is elliptically polarised. Then, if we consider a circle as a particular ellipse, we can characterise the flatness of the ellipse at the sample exit. To do this, the angle Ψ is determined from the calculation of its tangent, which involves dividing the smaller half-axis of the ellipse (A) by its larger half-axis (B) [96].
Since the ellipticity corresponds to a geometrical characterisation, thank to Figure 6.8 it can be seen that there is a maximum and a minimum value. The first is reached when $S = L$, that is when $\Psi = 45°$. In this case, the ellipse is in fact a circle, so the light is circularly polarized. The latter is reached when $S = 0$, i.e. when $\Psi = 0°$. This simply



(a)

$$\Psi = \arctan\left(\frac{A}{B}\right) \quad (6.3)$$

(b)

Figure 6.8: The ellipticity Ψ is (a) a geometric parameter determining the flatness of an ellipse, and is (b) calculated from the two half-axes of the ellipse.

corresponds to plane-polarised light. So we can say that linearly polarized light has an ellipticity of 0°.

However, the value of Ψ is usually very small, so it is expressed in milli-degrees (mdeg). This is why, one can consider that $\Psi = \tan(\Psi)$. Thus, another possibility to express the magnitude of the CD for this second parameter is to use the molar ellipticity $[\theta]$, expressed in $mdeg.cm^2/mol$:

$$[\theta] = \frac{\Psi}{C \times l} \quad (6.4)$$

Where C is the concentration [1] in $mol.L^{-1}$ and l is the optic path length in cm [96]. On the other hand, there is a variation of the molar ellipticity, applied to proteins and nucleic acids, it is the Mean residue molar ellipticity (MRME). Its calculation is based on the formula for molar ellipticity, but adds the number (n) of amino acid or nucleotide residues of the macromolecule of interest [104, 105].

$$MRME = \frac{\Psi}{C \times l \times n} \quad (6.5)$$

There is a relationship between the two parameters $\Delta\varepsilon$ and $[\Psi]$. Indeed, it has often been shown that for small values of Ψ, one can write the following equation [96, 106, 107] :

$$[\theta] = \frac{4500}{\pi} \times \log_e(10) \times \Delta\varepsilon \quad (6.6)$$

$$\approx 3298.2 \times \Delta\varepsilon \quad (6.7)$$

To finish, the CD spectrum is constructed by calculating the above parameters for each wavelength λ. For exemple, a spectra can be obtain by plotting $[\Psi]_\lambda$ in relation to λ.

[1] To remember $1L = 1000\,cm^3$. So the concentration in $mol.L^{-1}$ need to be divided by 1000 to convert it into $mol.cm^{-3}$ (equal to $mol.mL^{-1}$).

6.5 Origin of the G-quadruplex CD

Nucleic acids possess two main chromophores in the UV-Vis range: the ribose or deoxyribose sugar and the nitrogen bases. Purine and pyrimidine bases, have an electronic absorption at around 260 nm. But let's look more closely at the case of guanine, which is the characterizing base in G-quadruplexes. This base has two absorption bands in the UV-Vis region, involving two ($\pi \rightarrow \pi^*$) excited states at 248 nm and 279 nm [85], whose dipole moments are shown in Figure 6.9.

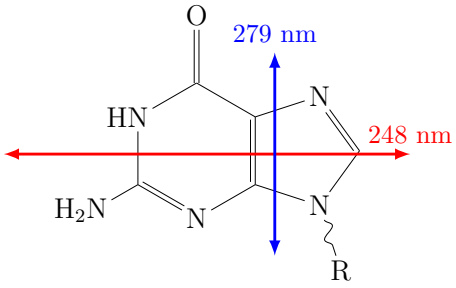


Figure 6.9: Representation of the two dipole transitions moments in guanine causing the relative electronic transitions. The most important is the one around 248 nm. *Adapted from Chaires and Graves [85].*

The other chromophore, the sugar, is characterized by an electrical transition around 180 nm. It also has a very interesting property in the context of circular dichroism, because it is an inherent chiral moiety that can assume D or L configuration (Figure 6.10) [108]. This property gives it the ability to provide a circular dichroism spectrum. However, the sugar chirality is usually not exploited in nucleic acid spectroscopic charcateristic since its absorption takes place in the far UV [96]. On the other hand, the secondary structure formed by nucleic acids procides a chiral arrangement, and hence induced a CD response. Specifically, the circular dichroism spectrum is derived from the electrical transition states of bases arranged in a helicoidal chiral structure. This effect is observed with double-helical DNA, as Z-DNA assumes a left-handed helix structure while both A-DNA and B-DNA assume a right-handed helix structure [109, 110, 111].

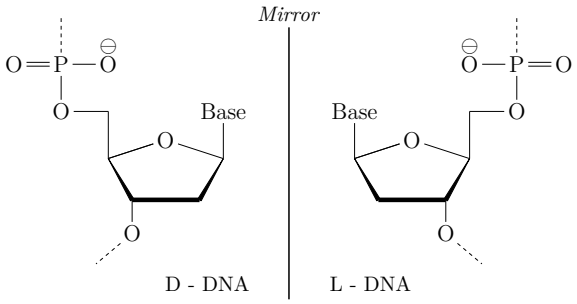


Figure 6.10: Charily of deoxyribose-phosphate sugar. *Adapted form [108].*

For similar reasons, also G-quadruplex structures adopt chiral right-handed or left-handed helical conformations (Figure 6.11) [85, 112]. However, this right-handed or left-handed helix folding does not seem to depend on the nucleotide sequence, but rather on environmental factors such as the presence of other molecules [113, 91, 114]. Interestingly, Winnerdy et al. [115] have shown experimentally that a G-quadruplex can adopt both types of helices forming what they name a " right- and left-handed hybrid G-quadruplex ". In conclusion, the electronic transitions of the guanines and their helical arrangement allows us to obtain a peculiar circular dichroism spectrum. As a matter of fact, CD signal is highly sensitive to the spoecific arrangements of supramolecular aggregates and hence it is a technique of choice for structural determination of nucleic acid arrangements.

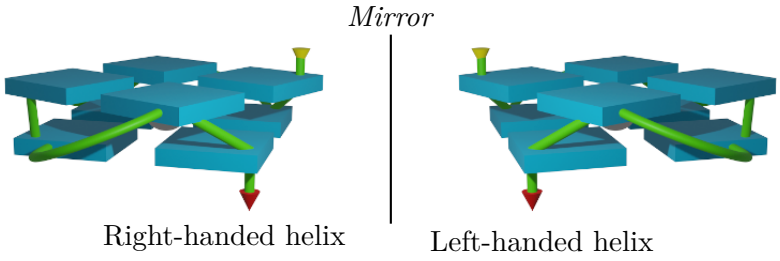


Figure 6.11: G-quadruplex is a chiral macromolecule.

6.6 G-quadruplex characterisation by CD

Circular dichroism spectroscopy is proving to be a very powerful technique for exploring G-quadruplex polymorphism. Each topology (arrangement of quartets and loops) has in fact its own spectrum. However, some generally applicable trends can be identified, as shown in Figure 6.12. For example, the spectrum of the parallel G-quadruplex is characterised by a negative peak at *ca.* 240 nm and a positive peak at *ca.* 260 nm. On the other hand, that of the hybrid G-quadruplex has an additional positive peak at *ca.* 290 nm. While the spectral signature of the antiparallel G-quadruplex differs from the previous two, showing a positive peak at *ca.* 240 nm, a negative peak at *ca.* 260 nm and a positive peak at *ca.* 290 nm [85, 78].

The spectral characteristics and their structural explanations have also been interpreted as in Table 6.1:

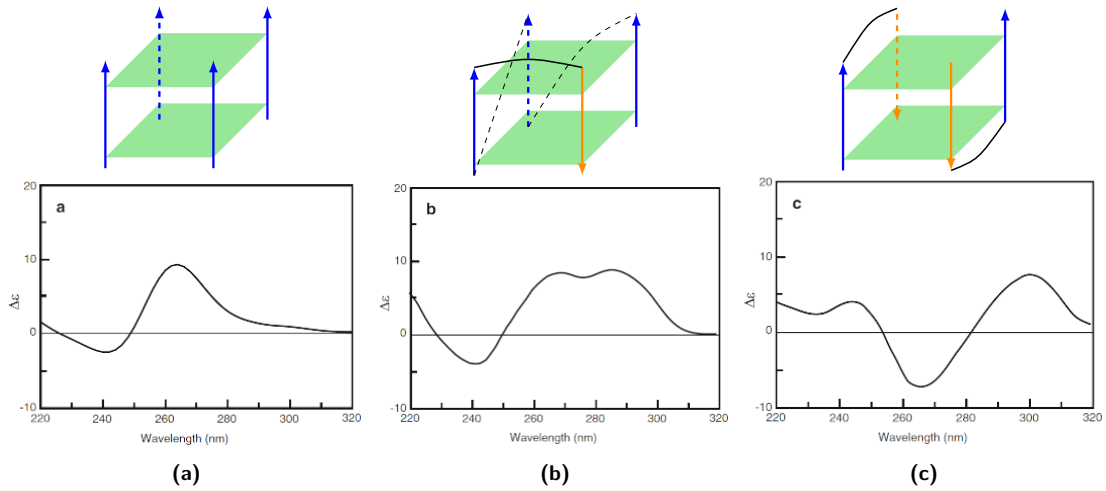


Figure 6.12: Examples of G-quadruplex CD spectra: **(a)** Tetrameric parallel G-quadruplex; **(b)** intramolecular hybrid G-quadruplex; **(c)** dimeric antiparallel G-quadruplex. *Modified from Chaires and Graves [85].*

Wavelength (nm)			Signification
240	*260*	*290*	
		+	Guanines stack with different glycosidic bond angles.
-	+		Guanines stack with the same glycosidic bond angles.
+	-		No guanines stacking with same glycosidic bond angles.

Table 6.1: Summary of G-quadruplexes circular dichroism spectra interpretation: + and - indicate positive and negative peaks, respectively. *Adapted from [85].*

CHAPTER 7

Structural parameters

Chapter contents

Daily practice shows that the development of various strategies for structural analysis of nucleic acids plays an important role in their classification. This started with the description of the helices of the different forms A-DNA, B-DNA and Z-DNA. Then the descriptions were refined to the point of describing the different conditions of the pairing between the bases [116, 117]. Today, software such as CURVES+ [118] or 3DNA [119] are able to provide information on the topology and base pairing of nucleic acids. However, even if the analysis of G-quadruplexes is possible, it is still limited to "general" topological

information. In fact, only few G-quadruplex-specific parameters are usually provided. However, several papers mention specific structural parameters for G-quadruplexes, we will focus on the article by Tsvetkov, Pozmogova, and Varizhuk [120], which gives several parameters specific to G-quadruplexes. Moreover, the authors provide a script for the VMD software [121] that allows to follow the evolution of those parameters along molecular dynamics trajectories.

Despite many efforts to define the structural parameters of G-quadruplexes, there is still a gap with respect to the description of the metal cation. Indeed, the totality of the parameters focus on the arrangement of the bases and the organisation of the sugar-phosphate chain. No method discusses the geometrical properties of the central cation with respect to the surrounding guanines. One exception is the article by Reshetnikov et al. [122], which uses the distance between the Centre of mass (COM) of the ion and the COM of the eight neighbouring O6 atoms.

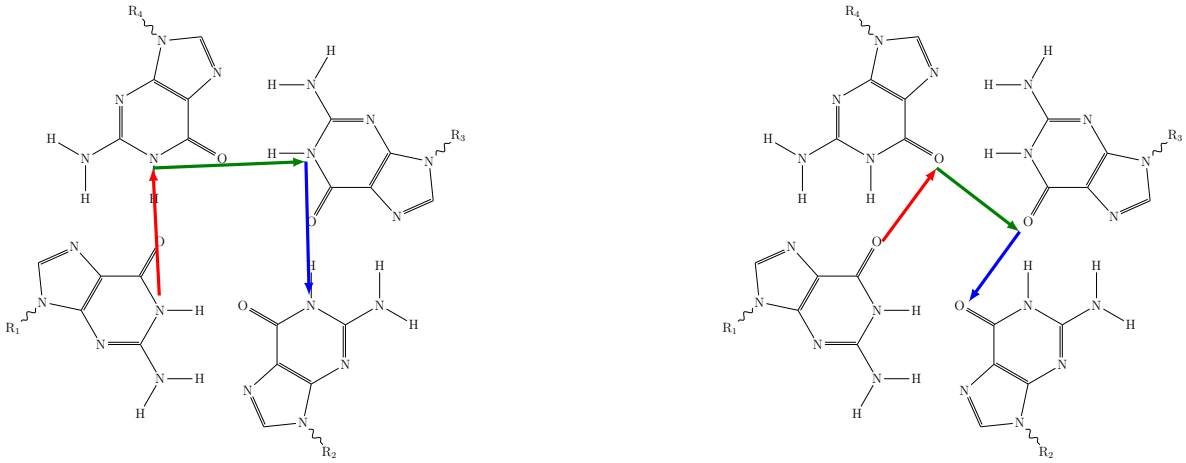
7.1 Angle parameters

7.1.1 Dihedral angle

A dihedral angle, also called a torsion angle, is constructed with three successive vectors that join a chain of four atoms A-B-C-D. More precisely, it is the angle formed between the plane formed by atoms A,B,C and that formed by atoms B,C,D [123]. The value of this type of angle is, in chemistry, a value between 0° and ±180°.

The dihedral angle is an indicator that highlights a out of plane bending of the G-tetrad. There are two specific variants. The strategy adopted by Mashimo et al. [124] is to construct a dihedral angle with the O6 atoms of each guanine involved in a quartet. The other strategy comes from Tsvetkov, Pozmogova, and Varizhuk [120] who propose to use the N1 atoms to construct the dihedral angle. Basically the two strategies are equivalent. The arrangement of guanines in a quartet gives a dihedral angle close to 0°. In fact, this value is never reached; it could only be reached if the four atoms were exactly coplanar. To better understand the construction of the dihedral angle within a G-tetrad, Figure 7.1 below represents graphically each of the two strategies. Later, we

will see other quartet bending parameters: "lengthwise" and "diagonal" quartet bending. Although they have their own characteristics, these three bending parameters give similar information.



(a) Dihedral angle involving four N1 atoms

(b) Dihedral angle involving four O6 atoms

Figure 7.1: Construction of the two dihedral angle variants in a guanine tetrad. In each representation, the dihedral angle is calculated between the red and blue vectors.

Mathematically the dihedral angle (Ξ) is calculated in the same way, whatever the type of atom, N1 or O6, is chosen. So the equation considers four atoms a, b, c and d as follows:

$$\Xi = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{abc}} \cdot \overrightarrow{n_{bcd}}\right) \tag{7.1}$$

Where $\overrightarrow{n_{abc}}$ and $\overrightarrow{n_{bcd}}$ are the normals of the planes abc and bcd, respectively. Both are calculated by the same equation, used here with $\overrightarrow{n_{abc}}$ as example :

$$\vec{n_{abc}} = \left[\frac{\vec{r_{ab}}}{\|\vec{r_{ab}}\|} \times \frac{\vec{r_{bc}}}{\|\vec{r_{bc}}\|}\right] \tag{7.2}$$

$\vec{r_{bc}}$ corresponds to the vector joining two adjacent N1 or O6 atoms.

7.1.2 Guanine-Guanine angle

As explained above, a quartet (or tetrad) is a group of four almost coplanar guanines, arranged in a square arrangement. Therefore, the angle between the longest side of two adjacent guanines should be approximately 90° and between two opposite guanines approximately 180°. If the angles between the guanines change drastically, then also the geometry of the quartet changes and can become an arbitrary convex quadrilateral. Thus

the guanine-guanine angle allows to monitor a geometric deformation of the quartet.

Now, it is appropriate to detail the calculation of this parameter. We will refer to Figure 7.2 to get a visual appreciation of the explanation given below. First, the axis of a guanine ($\overrightarrow{r^i}_{C4C5-C8}$) is defined by drawing a vector between C8 and the midpoint of the distance between the two C4 and C5 carbons. This is repeated for each guanine in the same quartet, to assign them an axis, then the angle between these axes is simply measured.

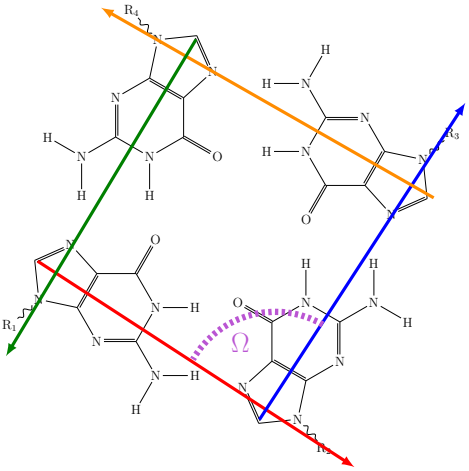


Figure 7.2: Ω is the angle between two guanine axis. In this example it's equal to 90° between red and blue vector, because they are adjacent. And it's equal to 180° between red and orange vector, because they are opposed.

Thus, the equation allowing to calculate the angle Ω may be written as:

$$\Omega_{i-j} = \frac{180}{\pi} \arccos\left(\frac{\overrightarrow{r^i}_{C4C5-C8}}{\left\| \overrightarrow{r^i}_{C4C5-C8} \right\|} \times \frac{\overrightarrow{r^j}_{C4C5-C8}}{\left\| \overrightarrow{r^j}_{C4C5-C8} \right\|} \right) \tag{7.3}$$

7.1.3 Normal of guanine - G-quadruplex Axis & Normal of two guanines angles

These two parameters allow us to define the orientation of a guanine with respect to the axis of the G-quadruplex, or to another guanine belonging to the same quartet. Together, they define a local planarity of the guanines in the tetrad, pointing out to some out of the plane deformations. This method requires a vector to be drawn between the centres of mass of two quartets called α and β, see Figure 7.3. In this way, we obtain an axis ($\overrightarrow{Z^{\alpha\beta}}$) which represents the vertical elongation of the G-quadruplex. But if the G-quadruplex has more than two quartets, such as three quartets, then the $\overrightarrow{Z^{\alpha\beta}}$ axis of

the firts quartet (α) is the vector joining the COMs of the α and β quartets, that of the second quartet $\overrightarrow{Z^{\beta\gamma}}$ is the vector joining the COMs of the β and γ quartets, so we consider that the axis of the γ quartet is the same as that of the β quartet which precedes it. Next, we need to establish the plane corresponding to the guanine. To do this, we draw two vectors: one connects the atoms N9 and N2 ($\overrightarrow{r^i}_{N9N2}$), while the other connects the atoms N9 and O6 ($\overrightarrow{r^i}_{N9O6}$). The two vectors thus created are tangent to the guanine, and form the plane corresponding to the guanine. Subsequently, the normal vector to the guanine plane $\overrightarrow{n_i}$ is drawn. This is a vector perpendicular to the tangent plane and passing through a point; it is calculated by the normalised vector product of the tangent vectors [125]. Tsvetkov, Pozmogova, and Varizhuk [120] provide the following equation:

$$\overrightarrow{n_i} = \left[\frac{\overrightarrow{r^i}_{N9N2}}{\left\| \overrightarrow{r^i}_{N9N2} \right\|} \times \frac{\overrightarrow{r^i}_{N9O6}}{\left\| \overrightarrow{r^i}_{N9O6} \right\|} \right] \tag{7.4}$$

Now the two essential elements are present, namely the vectors of the G-quadruplex axis and the normal vector on the guanine plane. Thus it is possible to calculate two parameters: the angle formed between the normal of the guanine and the G-quadruplex axis (φ), but also the angle formed between two normal vectors of two guanines ($\Delta\varphi$). The closer these two angles are to 0° the closer the guanine under study is to the plane of the quartet or the other guanine. They can be calculated using the following equations (Figure 7.3):

$$\varphi = \frac{180}{\pi} \arccos \left(\overrightarrow{n_i} \cdot \frac{\overrightarrow{Z^{\alpha\beta}}}{\left\| \overrightarrow{Z^{\alpha\beta}} \right\|} \right) \tag{7.5}$$

$$\Delta\varphi = \frac{180}{\pi} \arccos \left(\overrightarrow{n_i} \cdot \overrightarrow{n_j} \right) \tag{7.6}$$

This can also be seen graphically in Figure 7.3.

7.1.4 Twist angle

The helical twist angle, usually indicated as $\Theta^{\alpha\beta}$, represents the relative rotation angle of the G-quadruplex. More precisely, it determines the rotation of a quartet relative to the adjacent one. When extracted from of a molecular dynamics simulation, this parameter conveys information about the general flexibility of the guanine core. We have seen that G-quadruplexes are nucleic acid arrangements that can adopt a large

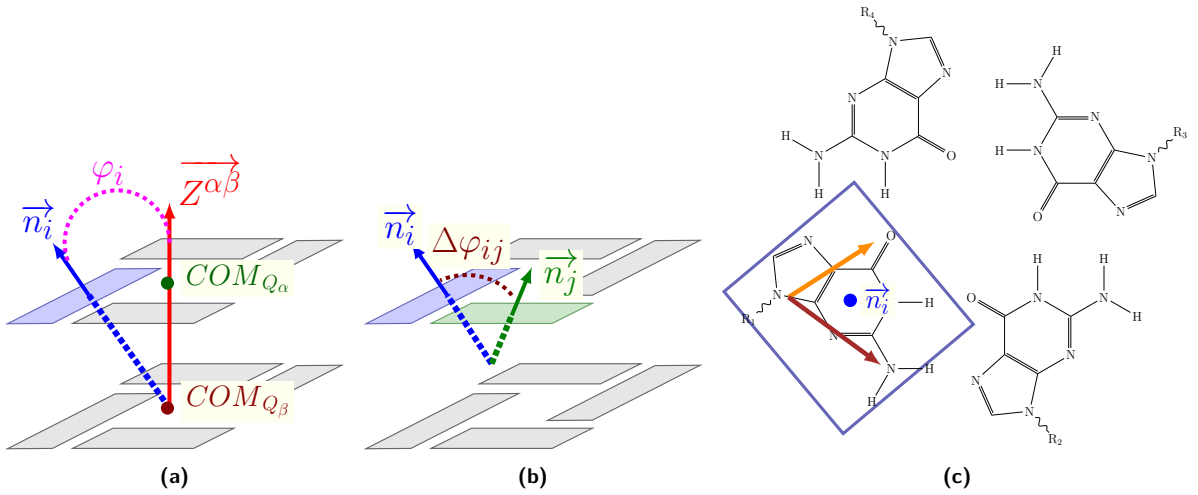


Figure 7.3: Plot of the construction of the angles φ (a) and $Delta\varphi$ (b), completed by the construction of the plane of a guanine and its normal (c).

variety of topologies. However, depending on its topology, a G-quadruplex has its own rigidity. That is, the greater the variation in the helical twist angle curve, the more flexible is the G-quadruplex. On the contrary, the smaller the variation, the stiffer the G-quadruplex.

The principle of measuring the twist angle is simple: measure the angle between two quartets. Lech, Heddi, and Phan [92] establish the value reference, i.e. 0, value of this angle according to the stacking of the O6 atoms. Indeed, the authors have arbitrarily chosen that the value 0° corresponds to the smallest separation distance of the O6 atoms of the stacked guanines. This is the arrangement illustrated their original contribution and reproduced here in Figure 7.4. This method is interesting and quite innovative. Nevertheless, it is not based on structural evidence. In fact, it requires a pre-established knowledge of the stacking of the quartets. Secondly, it makes difficult to compare the twist angle between two G-quadruplexes characterized by a different topology.

Other methods of measuring the helical twist angle have been reported. One of them also uses the O6 atoms of guanines. For each tetrad, these atoms globally form a planar square. Thus, it is sufficient to measure the angle formed between the squares of each quartet [78]. Reshetnikov et al. [126] proposes another method based on the use of the C1' atoms of the guanosine sugar. To do this, we draw a vector between the C1' atoms of two adjacent guanines in the first quartet. Then we do the same in the next quartet. In this way, we obtain two vectors, one for each quartet. The measure of the angle between these two vectors determines the twist angle. As a matter of fact, the method proposed

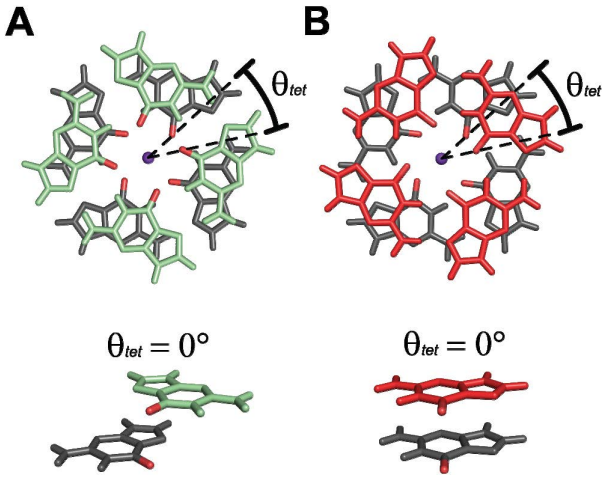


Figure 7.4: Illustration of the method established by Lech, Heddi, and Phan [92] to define the 0° of the twist angle for quartets with opposite polarity (a) or having the same polarity (b). *Figure reproduced from Lech, Heddi, and Phan [92].*

by Tsvetkov, Pozmogova, and Varizhuk [120] is equivalent to the previous one, but uses N1 atoms, as shown in Figure 7.5.

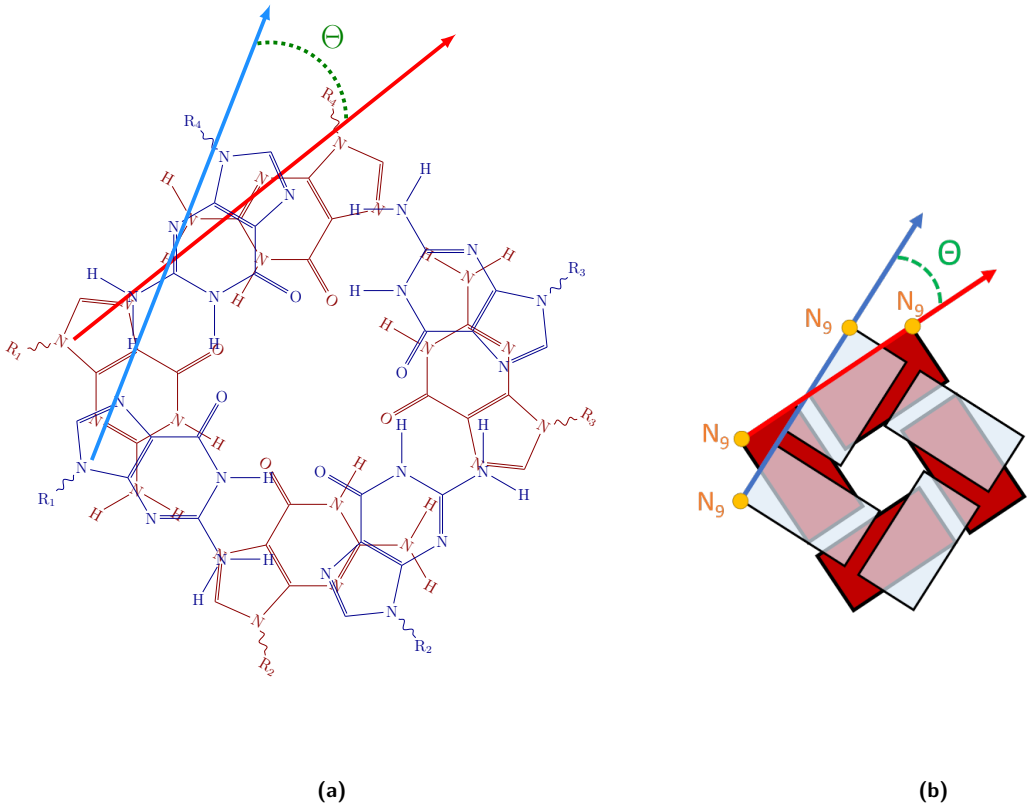


Figure 7.5: Detailed (a) and simplified (b) representation of the method used by Tsvetkov, Pozmogova, and Varizhuk [120] to measure the helical twist angle (Θ) of G-quadruplexes. The angle is measured between the vectors connecting the N1 atoms of adjacent H-bonded guanines.

Contrary to the first one, this protocols introduce less arbitrary choices and are grounded on structural evidences, which allows them to compare the twist angle of G-quadruplexes independently of their quartet stacking, while they can still be easily

implemented in a software. So, the angle between the vectors of the α quartet and the β quartet is calculated using the following equation, where $\Theta_{ij}^{\alpha\beta}$ is the angle, $\overrightarrow{r_{ij}^{\alpha}}$ is the vector that join guanines i and j of the α quartet by the N1 atoms. $\overrightarrow{r_{ij}^{\beta}}$ is the same angle for the β quartet.

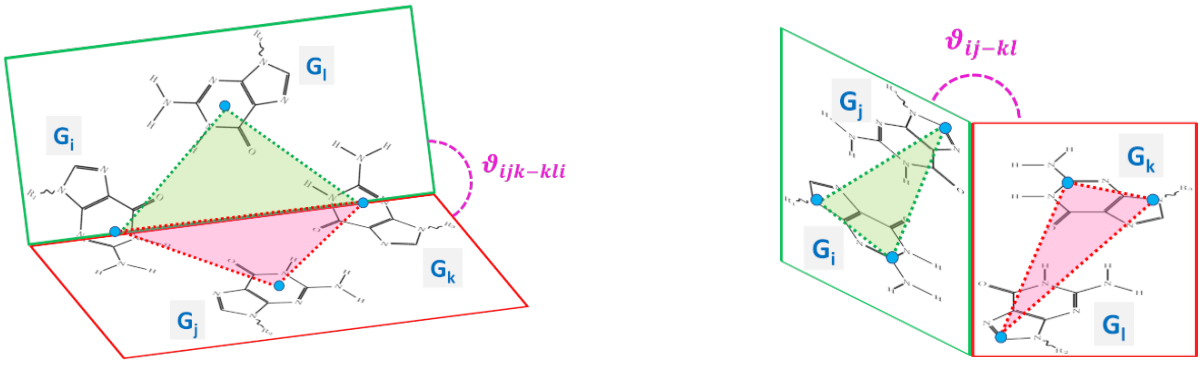
$$\Theta_{ij}^{\alpha\beta} = \frac{180}{\pi}\left(\frac{\overrightarrow{r_{ij}^{\alpha}}}{\left\|\overrightarrow{r_{ij}^{\alpha}}\right\|} \times \frac{\overrightarrow{r_{ij}^{\beta}}}{\left\|\overrightarrow{r_{ij}^{\beta}}\right\|}\right) \tag{7.7}$$

Indeed, this formula is not complete, because it does not refer to all guanines of the tetrad. So the formula of the twist angle is the arithmetic average of the four angles formed by all the guanines, namely:

$$\Theta^{\alpha\beta} = \frac{\Theta_{ij}^{\alpha\beta} + \Theta_{jk}^{\alpha\beta} + \Theta_{kl}^{\alpha\beta} + \Theta_{li}^{\alpha\beta}}{4} \tag{7.8}$$

7.1.5 "Diagonal" and "Lengthwise" bending angles

These two parameters both express the bending of a quartet. They give a more global view of the information provided by the Normal of Guanine - G-quadruplex Axis angle. They are however defined in the same way as the previous ones: the quartet is cut into two planes and the angle formed between them is measured. Then we draw the normal of each plane and measure the angle form by these two vectors.



(a) "Diagonal" bending

(b) "Lengthwise" bending

Figure 7.6: Representation of the construction of the guanine planes and the measurement of their angle, according to the two bending parameters. *Adapted from Tsvetkov, Pozmogova, and Varizhuk [120].*

The "Diagonal" bending parameter ($\vartheta_{ijk-kli}$) uses the center of mass of three guanines to form a plane. Consequently the two planes correspond to those formed by the triangles

connecting the centers of mass of guanines *i,j,k* and *k,l,i*. The normal of each plate is calculated by an identical formula to the one expressed here for the *i,j,k* plane :

$$\overrightarrow{n_{ijk}} = \left[\frac{\overrightarrow{n_{ij}}}{\|\overrightarrow{n_{ij}}\|} \times \frac{\overrightarrow{n_{jk}}}{\|\overrightarrow{n_{jk}}\|} \right] \tag{7.9}$$

Thus, the "Diagonal" bending angle is:

$$\vartheta_{ijk-kli} = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{ijk}} \cdot \overrightarrow{n_{kli}}\right) \tag{7.10}$$

The "Lengthwise" bending parameter (ϑ_{ij-kl}) uses two vectors to define each plane. The first vector connects the N9 and C2 atoms of guanine *i*. While the second vector connects the N9 atom of guanine *i* to the C8 atom of guanine *j*. The *ij* plane is thus created; the *kl* plane is drawn in a similar way. Then the normal of both planes are calculated in the same way, detailed here for the *ij* plane:

$$\overrightarrow{n_{ij}} = \left[\frac{\overrightarrow{n^{i}_{N9C2}}}{\left\|\overrightarrow{n^{i}_{N9C2}}\right\|} \times \frac{\overrightarrow{n^{ij}_{N9C8}}}{\left\|\overrightarrow{n^{ij}_{N9C8}}\right\|} \right] \tag{7.11}$$

Finally, the "Lengthwise" bending angle is calculated following the equation:

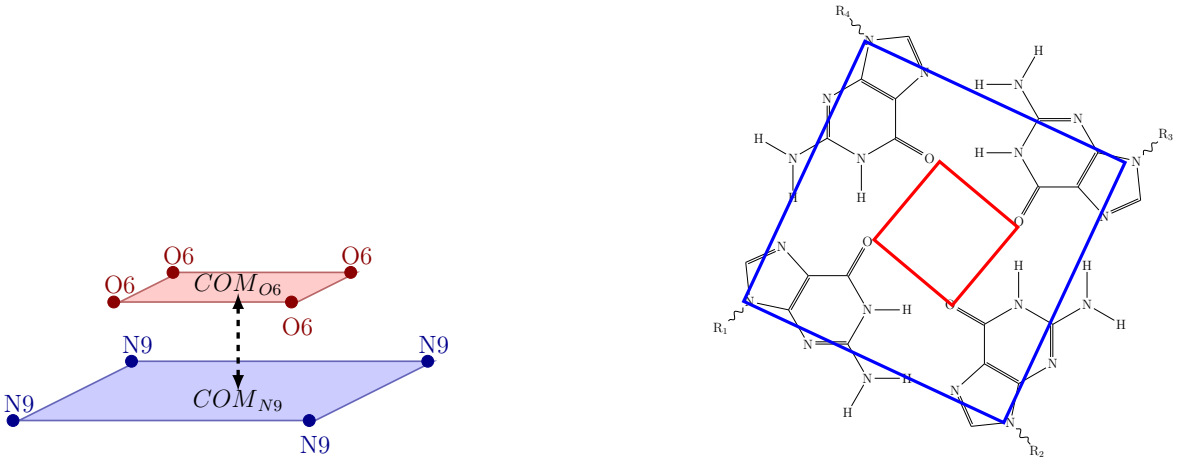
$$\vartheta_{ij-kl} = \frac{180}{\pi} \arccos\left(\overrightarrow{n_{ij}} \cdot \overrightarrow{n_{kl}}\right) \tag{7.12}$$

7.2 Distance and area parameters

7.2.1 Planarity of quartets

Planarity (ρ) corresponds to the fact that all the guanines of a quartet are coplanar. The distortion from planarity can be easily calculated by the distance between the center of mass of the tetragon formed by the four O6 atoms and the center of mass of the tetragon formed by the four N9 atoms [127], see Figure 7.7, using this equation:

$$\rho = COM(N9^{tetragon}) - COM(O6^{tetragon}) \tag{7.13}$$



(a) Separation of the O6 and N9 tetragons

(b) Construction of the O6 and N9 tetragons

Figure 7.7: Graphical representation of the calculation of the planarity of a quartet. *Adapted from Reshetnikov, Golovin, and Kopylov [127].*

7.2.2 Separation of G-tetrads

The separation distance d_{tet} is the distance between parallel quartets [92] in a G-quadruplex. It is obvious that tideally such a definition implies that the guanines in the quartet should be coplanar, see Figure 7.8. This parameter is equivalent to the separation distance of the stacked base pairs in double stranded DNA.

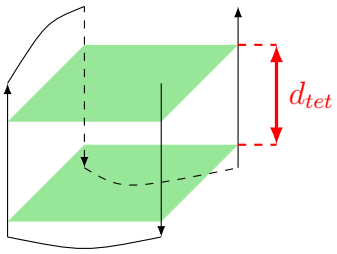


Figure 7.8: Schematic representation of the separation distance between G-tetrads.

The separation distance is calculated by the difference of the center of mass of each quartet stacked on the previous one along the axial elongation of the G-quadruplex:

$$d_{tet} = COM(quartet^{\alpha}) - COM(quartet^{\beta}) \tag{7.14}$$

7.2.3 Guanine-Quartet COMs and Guanine-G4 COMs distances

These are two very useful parameters because they describe the extrusion of each guanine from the quartet or from the G-quadruplex. They complement the planarity parameters in the description of G-quadruplex structural conservation and rigidity.

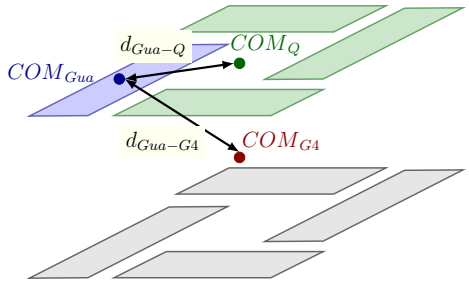


Figure 7.9: d_{Gua-Q} and d_{Gua-G4} are established according to the difference of two mass centers.

As highlighted in Figure 7.9, each parameter is a centre of mass distance. The first is the distance between the centre of mass of the guanine under study and the centre of mass of its quartet (Q). The second is the distance between the centre of mass of the guanine and that of the whole G-quadruplex. Thus we obtain the following equation:

$$d_{Gua-Q} = COM(guanine) - COM(quartet) \tag{7.15}$$

$$d_{Gua-G4} = COM(guanine) - COM(G-quadruplex) \tag{7.16}$$

7.2.4 Area of the O6 tetragon

Reshetnikov et al. [122] use this parameter to track the space between the four guanines in a tetrad. To do this, the four O6 atoms are linked to form a tetragon. Monitoring the evolution of its area makes it possible to follow any dynamical deformation of the guanine tetragon.

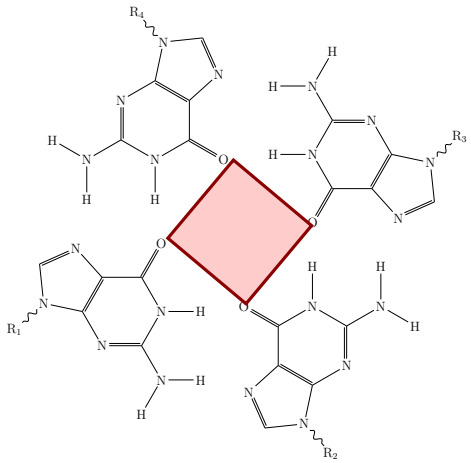


Figure 7.10: Plot the area of the tetragon formed by the four O6 atoms.

7.3 Compactness of G-quadruplexes

The radius of gyration is defined as the root-mean-square average of the distance of all scattering elements from the center of mass of the molecule [128]. It highlights the compactness of a molecule along a molecular dynamics simulation, because the smaller the value, the closer the atoms of the molecule remain relative to the global centre of mass. In fact, its monitoring in different conditions allows to see if an element favours the compaction, or decompaction, of a G-quadruplex. However, the radius of gyration while providing a description of the maintaining of the folding pattern of a given macrostructure, applies preferentially to globular entities. Figure 7.11 underlines the concept presented here, by comparing the radius of the sphere surrounding the molecule with the radius of gyration, which is smaller than the sphere surrounding the molecule.

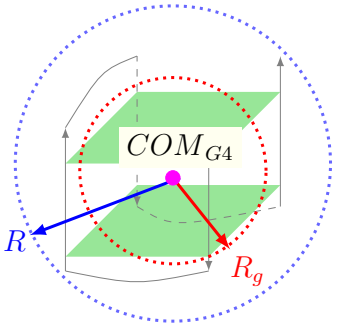


Figure 7.11: The difference between the radius of the sphere (R) enclosing the G-quadruplex and its radius of gyration (R_g).

3

Computational methodology

Simulation methods as defined by computational chemistry are able to provide information that may be experimentally difficult to obtain. This may be due to practical reasons: the difficulty to obtain a crystal, or it may be due to other methodological limitations such as systems too large to be analyzed by NMR. In the work presented here, simulations are used to answer a structural question centered around the impact of damages introduced at specific locations in G-quadruplexes, to understand the structural mechanisms of the interaction of its non-canonical forms with proteins, but also to expose a way to predict the folding type of a G-quadruplex. These objectives were achieved using different computational methods: classical and QM/MM molecular dynamics simulations and DFT calculations.

Molecular dynamics is a method allowing to simulate the temporal evolution of the structure or the reactivity of a molecular system. To do this, it uses models of molecules that are sufficiently complete to be useful, but sufficiently simple to be computable. Two main approaches limit the field of application of theoretical chemistry:

- Simulations using classical molecular mechanics. The equations of motion are numerically integrated for the nuclei of the atoms of the studied system. It relies on the use of *ad hoc* parameters, which depend on the system studied: geometry of the molecule and types of atoms. More precisely, molecular mechanics uses a set of potential functions that describe the energy of the system: these are called the force fields. This type of approach allows the calculation of trajectories of several thousand atoms over long times, of the order of microseconds (μs). Thus this type of simulation is used for the structural analysis of molecules. For example, its use allows to study the interaction between a protein and a ligand, or to observe the structural change of a protein or a nucleic acid.

- Simulations based on the principles of quantum physics: *ab initio*. It focuses on the electronic properties of the molecular system, allowing to simulate the formation and breaking of chemical bonds. Contrary to classic molecular mechanics, this type of simulation does not use any parameters and seeks to solve the Schrödinger equation for the molecular system. The absence of parameterization makes the method applicable to any type of molecular system. However, the use of this approach makes the calculations more expensive and the method is limited to a hundred atoms and to shorter simulation times. For example, this type of simulation is

used for the exploration of chemical reactivity that may take place in an enzyme.

Two other methods are intermediate between the classical method and the (purely) *ab initio* method. These are methods that integrate a certain degree of parameterization: fixed parameters are implemented in the equations to simplify the quantum representation of the molecular system. Consequently, the use of a parameterization lightens the calculations and gives access to longer times. The DFT method can use parametrization and is close to the *ab initio* method. On the contrary, the semi-empirical method diverges a little from the *ab initio* and is closer to the classical approach, while still giving access to chemical reactivity. Because it uses a parametrization depending on the geometry and the type of atom [129].

Table 7.1 below summarizes the properties of the four main methods presented in these paragraphs. But the different aspects of these methods are detailed in the following chapters.

Methods / **Features**	***ab initio***	**DFT**	**semi-empiric**	**empirical (molecular mechanics)**
Period of emergence of the method	1950's (resurgence with first computers)	1990's	1970's	1960's
First popular software using the method	GAUSSIAN 70 (1970)	ADF (1995)	MOPAC (1990)	MM2 (1977)
Theoretical formulation	Quantum (without parametrization)	Quantum (with parametrization)	Quantum (with chemical element by chemical element parameterization)	Classical (parametrization by type of atoms)
Modelled object	Small molecules	Large molecules	Large molecules	Macromolecules
Performance criteria	Representation in adequacy with quantum theory	Electronic prediction	Thermodynamic prediction	Conformational prediction

Table 7.1: Presentation of the characteristics of the four main methods used in theoretical chemistry. *Reproduced from Bensaude-Vincent and Eastes [129], personnal translation.*

CHAPTER 8

Classical molecular dynamics

I can calculate the movement of heavy bodies, but not the madness of crowds.

Sir Isaac Newton

Chapter contents

8.1 Validity of the classical treatment

To simulate a molecular system, it would be more rigorous to solve the Schrödinger equation corresponding to the system under study. But, it is impossible to perform this type of calculation on large systems, like proteins. The solution is to use Newton's law equation to simulate the motion of each nuclei on top of an empirical parametric potential. The evolution of the position of the nuclei over time gives a trajectory. Despite its approximation, this approach is justified by two important properties. First, the Born-Oppenheimer approximation, proposed in 1927 [130], states to the fact that the mass of the nuclei of atoms is much greater than that of the electrons. This makes it possible to consider the motion of the electrons and the motion of the nuclei separately. Based on this principle, molecular dynamics takes into account only the nuclei of atoms. Then, in most chemical systems, the De Broglie wavelength, recalled below, is smaller than the distance that separates an atom from its nearest neighbor. Therefore, nuclear quantum effects are essentially negligible [131, 132]. So, provided that the potential energy is well represented by the empirical potential, classical molecular dynamic is valid for structural evolution, but gives no information over bond breaking and chemical reactivity or excited states.

$$\lambda = \sqrt{\frac{h^2}{Mk_BT}} \tag{8.1}$$

Where M is the atomic mass and T correspond to the temperature, k_B is the Boltzmann constant and h symbolizes the Planck constant.

8.2 The phase space

The equations of molecular dynamics are described in phase space [133]. That is to say, an abstract space with *2f* dimensions; in which f represents the number of degrees of freedom. More precisely, this space results from the set of physical quantities required to calculate the evolution of the studied system. In the framework of classical dynamics presented here, the physical quantities are the position and the corresponding

momentum; the phase space is therefore the space of states of the system. It is important to know the values of these quantities at a precise time space t. In this sense, any point in the phase space is specific to a state of the system under study [134].

Stöcker et al. [134] explains that a trajectory which is carried out in the space of phase re-situates the periodic movements which give the temporal evolution of the studied system. This means that the evolution of the system in the course of time is dependent on the position of the system in the phase space at a given instant t. But also that different trajectories cannot cross each other in the phase space. Otherwise it would not be possible to know with precision what is the trajectory followed by the system to arrive at this point.

We recall here that each atom is localized in space by 3 Cartesian coordinates (x,y,z). So, to describe a molecule of N atoms $3N$ coordinates, corresponding to $3N$ degrees of freedom, are necessary. Among these, 3 degrees of freedom correspond to rotational motion and 3 to translational motion of the whole molecule. Such motions are not involved in the vibration of the molecule. Consequently, the vibration modes of a molecule, involving changes in the relative distance and the angle between atoms, are $3N - 6$ (or $3N - 5$ for linear molecules) [131].

8.3 Equations of classical molecular dynamics

8.3.1 Newton's equations

Hamilton's equations are used to describe the motions of each atoms (i) of a molecule along a trajectory.

$$\begin{cases} r_i = \dfrac{dH}{dp_i} \\ q_i = -\dfrac{dH}{dr_i} \end{cases} \tag{8.2}$$

The equations involve the momentum (p_i) and the position of the atom (r_i). But they can be reduced to the Newtonian equation of motion:

$$f_i = m_i \times a_i \tag{8.3}$$

In which the acceleration (a_i) is the derivative of the velocity (v_i), which is itself the derivative of the position r_i. So the acceleration is the second derivative of the position. Hence the following statement, where E is the energy of the system with respect to the position of the atom i :

$$\begin{cases} f_i = m_i \dfrac{d^2 r_i}{dt^2} \\ f_i = -\dfrac{dE}{dr_i} \end{cases} \tag{8.4}$$

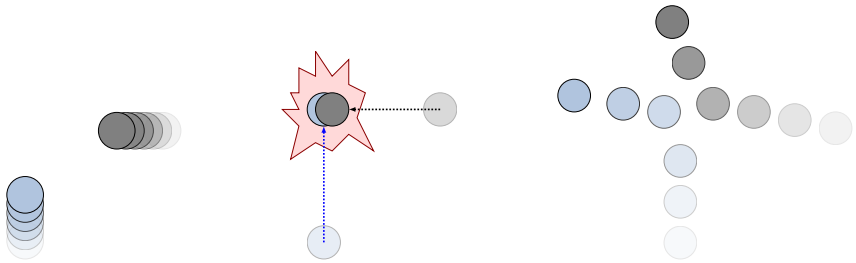
So a molecular dynamics trajectory is performed by the integration of Newton's equations, for which the force (f_i) emanates from the potential defined by the force field. [133].

To finish, it is important to note that in real systems the equations cannot be solved analytically so we need to integrate them numerically. This leads to the definition of a time step.

8.3.2 The integration step

The numerical integration of the equation of motion implies the definition of a discrete time advancement according to a time step (δt). The time step is a crucial parameter for the stability of the dynamics. More precisely, a correct time step must guarantee the conservation of the energy of the system [133, 131]. If the chosen value is too small then the overall computational cost will increase so much that the simulated time will be too short to obtain a correct trajectory, despite a more faithful approximation to the differential equations. But if the chosen value is too large, then there is a considerable increase in the energy of the system. In such a situation, the system is unstable and it is not possible to obtain a physically meaningful trajectory [133, 131]. Figure 8.1 explains these principles in a more visual way. In concrete terms, the choice of time step is made according to the movements with the highest frequency in the system. In fact, it is important that the time step is smaller than the period of these motions: $\delta t \ll v_{movements\ max}$. In molecular dynamics the motions with the highest frequencies are the vibrations of covalent bonds involving hydrogens. The choice of a time step of 1 fs

is therefore recommended [131, 132].



(a) δt too low **(b)** δt too big **(c)** δt suitable

Figure 8.1: Effect of poor and good integration step selections. *Modified from Fuchs [135].*

The heaviness of molecular dynamics computations often tests the computational speed of computers. The need to respond to the decrease in computation time has led to the development of several approaches whose goal is to increase the time step. We can mention the SHAKE [136] and SEETLE [137] algorithms. SHAKE applies constraints by adding a force on the covalent bonds of hydrogens, for example on a bond of type C — H. This action allows the use of a time step twice as high: it goes from 1 fs to 2 fs. SETTLE instead fixes the geometry of the solvent, water, molecules. Nevertheless, the use of constraints is not the only method to increase the time step. In fact these algorithms can be coupled to the Hydrogen Mass Repartitioning (HMR) and thus increase the time step up to 4 fs. This approach consists in increasing the mass of all the hydrogens, except those of water molecules, from 1.008 to 3.024 a.u. However, the total mass of the system is conserved because the mass added to the hydrogens is removed from the heavy atoms directly bound to them as show in 8.2 Indeed, the increase of the hydrogen mass will induce a slowing down of the higher frequency movements.

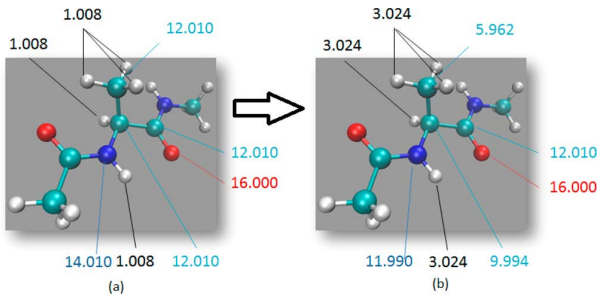


Figure 8.2: Example of the application of an HMR on a peptide. (a) Canonical dialanine peptide atomic masses (b) and after Hydrogen Mass Repartitioning *Reproduced from the article of Hopkins et al. [138]*

8.3.3 Cut-off

The application of the above equations does not require great computational complexity if the system is very small. But most of the time the size of the studied systems

requires to restrict the number of interactions remaining in the system. Doing so allows to reduce the computation time and the memory [131, 132]. The method needed here is called a cut-off. It consists in ignoring the interactions located at an arbitrary distance around a particle i. The application is quite simple. First we build the list of neighbors of the particle, using the $R_{neighbors}$ distance. This list is updated regularly, and thanks to it all the atoms that are located in the cut-off radius are established. Obviously, the cut-off radius R_{cutoff} is smaller than the one used to establish the list of neighbors (see Figure 8.3) [133, 131].

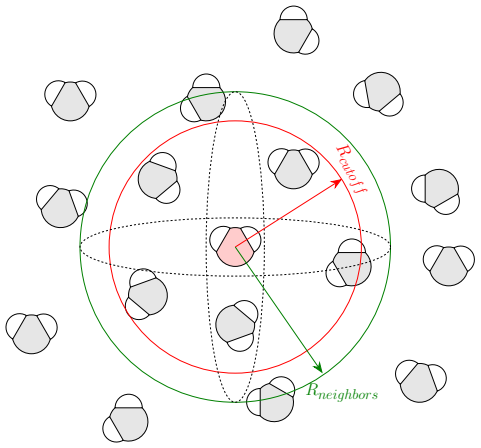


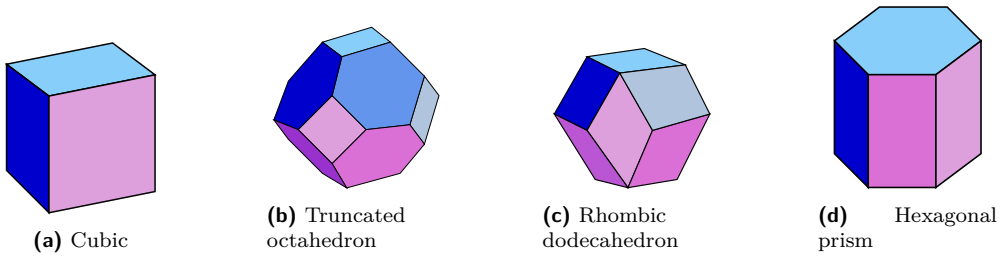
Figure 8.3: The particle i is an oxygen of a water molecule marked in red. The list of its neighbors is established with the $R_{neighbors}$ radius. Then the cutt-off R_{cutoff} is used to calculate the interaction between the particles inside this sphere.

8.4 Solvent, solvation and periodic boundary conditions

8.4.1 Using water as solvent

Water constitutes about 70% of the cells [139]; it is the solvent of biological systems. This is why the simulated molecules are solvated in water. In other words, water molecules are added all around the molecule to be simulated. The set is called a box and it has the shape of a geometrical polyhedron (Figure 8.4); the most common forms are the cube and the truncated octahedron. During the calculations, water is described by a force field parameterized especially for reproducing the properties of this liquid solvent. Some of the most used potentials for water are the TIP3P model (Transferable Intermolecular Potential 3 Points) or the SPC model (Single Point Charge). Bonomi and Camilloni [140] explains that these two models are very similar in that the water

molecule is described by three harmonic potentials, namely: the O——H bond, the H——H bond and the valence angle $\widehat{HOH}$ established between the three atoms.



(a) Cubic **(b)** Truncated octahedron **(c)** Rhombic dodecahedron **(d)** Hexagonal prism

Figure 8.4: The simulation boxes are polyhedra.

8.4.2 Periodic boundary conditions

The use of a simulation box implies the definition of a finite, and usually relatively small, system. However, the use of this type of object is not rigorous as regards the reproduction of thermodynamic quantities and produces edge effects. The solution is to virtually replicate the box in all directions, which produces a pseudo infinite system [133, 131]. However, there is a drastically increase of the interactions in the new system. Consequently, the use of a cut-off becomes necessary to limit the calculation overload. This means that a particle i in the minimal box is able to interact with another particle j in the minimal box, but also with a particle j' in a virtual box, if they are within the cut-off limit [133]. The use of cut-off also avoids unphysical self-interaction of the studied solute with its own image in a nearby box.

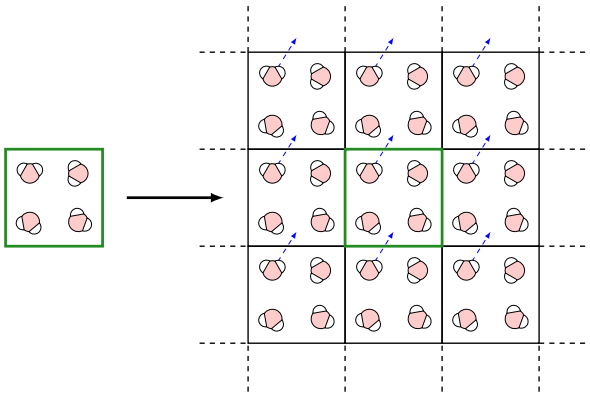
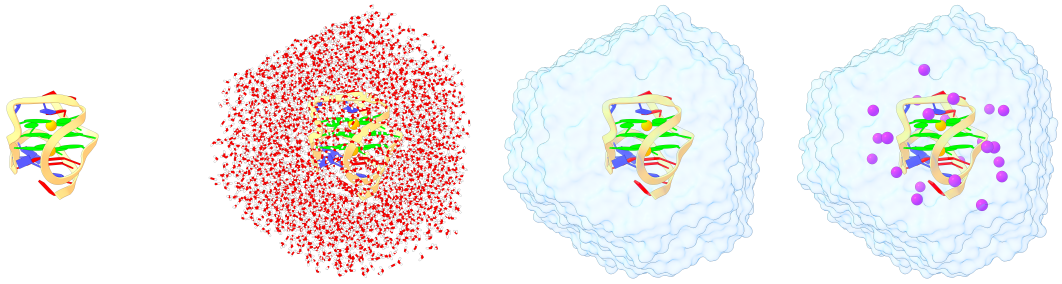


Figure 8.5: Periodic boundary condition of a simulation box.

8.4.3 Solvation and electrical neutrality

Molecular systems are electrically neutral, however this is not always the case with the solvated molecule. In this sense, it is necessary to ensure the electrical neutrality of the calculated system because of periodic boundary conditions. Electrical neutrality is ensured by the addition of ions: some water molecules are replaced by ions, until electrical neutrality is reached. The process of solvation and addition of ions is schematized by the Figure 8.6. Secondly, it goes without saying that the addition of ions is not exclusively limited to charge neutralization. Indeed it is possible to add cations and anions to obtain a certain salt concentration in the simulation box, namely the physiological concentration of 0.15 M.



(a) Molecule preparation **(b)** Solvatation **(c)** **(d)** Addition of ions

Figure 8.6: During the solvation process, water molecules are added around the molecule, while respecting the imposed geometrical shape; then the ions are added.

8.5 Statistical sets

The statistical sets represent thermodynamic constraints imposed by the external environment on the system [131, 132]. The thermodynamic parameters found in the canonical sets are: the number of atoms (N), the pressure (P), the temperature (T), the energy of the system (E), the volume of the box (V), and the chemical potential (μ). Canonical sets are intended to keep some of these parameters constant [131, 140]. Now we can expose the four most frequently encountered sets:

- **NVE** : Here the constant parameters are the number of atoms (N), the volume of the box (V) and the total energy of the system (E). Solving Newton's equation of motion in PBC gives NVE because the equations conserve total energy [131].

- **NVT** : In this set the constraint of constancy applies to the number of atoms (N), the volume (V) and the temperature (T) [131]. To obtain the NVT, it is necessary to apply a thermostat. To obtain the NVT, it is necessary to apply a thermostat. This means that the velocities must be sclaed to maintain the kinetic energy, and thus hence T constant.

- **NPT** : [131, 132] explains that this set is much more representative of the experimental conditions. In this case, the constraint of constancy is applied to the number of atoms (N), the pressure (P) and the the temperature (T). To obtain the NPT, it is necessary to use a barostat. This means that the volume is changed to keep the pressure constant.

CHAPTER 9

Parameterization of molecular force fields

Chapter contents

9.1 What is a molecular force field ?

Force field is a term that has become widely used, also in science fiction and fantasy movies, books and video games. In the latter context, its meaning is defined as a barrier of energy (physical or magical) serving as a passive protection system, such as a shield, intended to repel various forms of danger, or serving as a trap to enclose enemies [141].

In physics, the notion of force field is a portion of space-time in which a constant force is exerted and acts outside of contact on an object located in this field. The magnetic field or the gravitational fields are two examples.

In computational chemistry, the definition of force field is slightly different from that used in physics. In the context that interests us here: a force field is a function that uses the positions of the atoms to describe the variations of the potential energy surface of the molecule [140, 131].

9.2 Overview of empirical potential energy calculations

In molecular dynamics, a molecule is considered as a set of atoms linked together by chemical bonds. Each atom is described as a ball with its own radius, mass and partial charge (or a complete charge for monoatomic ions). Generally, we distinguish two types of interatomic interaction energies: the one between bound atoms and the one between unbound atoms, including also atoms of different molecules (intermolecular interactions). Thus the empirical potential energy ($E(r^N)$) for N particles system, depending on the Cartesian coordinates r_i, is the sum of these two terms [133, 131, 140].

$$E(r^N) = E_{bonded\ atoms} + E_{unbonded\ atoms} \tag{9.1}$$

This description is very minimalist. In fact, each of the two energies is itself the sum of several terms. The energy of bound atoms takes into account the elongation energy of

covalent bonds, as well as bond angle and dihedral angle strain energies. The interaction energy of unbound atoms includes the Van der Waals and the electrostatic components (the latter is also called Coulomb energy) [131, 133]. All these terms are explained below. Before continuing, it is important to understand that the terms of a potential energy function do not always have a straightforward physical or chemical meaning, instead they are meant to reproduce as much as possible the form of the potential energy surface. Some terms have been simply calibrated [133, 140] and their equilibrium values are obtained by comparison with data obtained by experimental measurements or by high level quantum chemical calculations.

9.3 Interaction between bound atoms

9.3.1 Elongation energy of covalent bonds

The elongation term is a parameter that establishes the energy potential necessary to stretch a covalent bond between two atoms. The bond oscillates around an equilibrium value for which the potential energy of the two atoms is minimal (called E_0) [142], see Figure 9.1. The oscillation always takes place in an acceptable zone, i.e. close to the equilibrium value, so that the values of distance (r) and equilibrium distance (r_0) are sufficiently close. The parameter E_{bond} is an analytical function which expresses the variation of the value of the bond distance around its equilibrium value, using a force constant of the bond (k_r). The most used equation is a harmonic function, which comes directly from the second order truncation of the Taylor expansion of the potential. Harmonic approximation is simple and requires a relatively small computation time [133].

$$E_{\text{bond}} = k_r \left(r - r_0\right)^2 \tag{9.2}$$

However, the harmonic function does not perfectly describe the stretching-compression of a chemical bond. In fact, the variation of energy as a function of distance, especially for large deviations from the equilibrium, is more accurately reproduced by the Morse dissociate potential function. We can visually compare the harmonic and the Morse functions in Figure 9.1. The Morse function should be used when theoretical research requires a more accurate level of calculation [133, 131]. But only harmonic terms are

used in the methodology of this work.

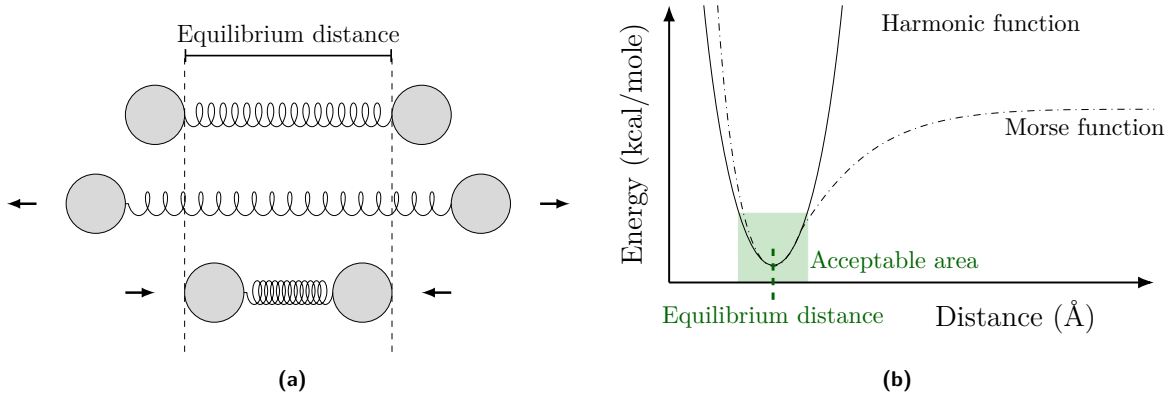


Figure 9.1: **(a)** A covalent bond can be considered as a sprong and the bonded atoms oscillate around an equilibrium distance, corresponding to the E_0 energy value, by attraction-repulsion interactions. **(b)** Comparison between the harmonic and the Morse functions. The harmonic (quadratic) function describes well the bond stretching for small deformations around the equilibrium value.

9.3.2 Angular strain energy

The (E_{angle}) energy describes the bending of the bond angle (θ) around an equilibrium value (θ_0), as shown in Figure 9.2. The relative equation describing this phenomenon is typically a harmonic function that uses a stiffness constant (k_θ). However, also in this case the bond angle bending energy can be improved by using cubic and higher order corrections, as seen for the stretching energy of covalent bonds [131, 132].

$$E_{\text{angle}} = k_\theta \left(\theta - \theta_0\right)^2 \tag{9.3}$$

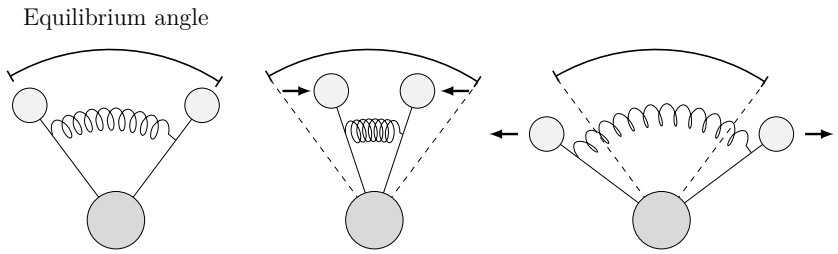


Figure 9.2: The bond angle oscillates around an equilibrium value.

9.3.3 Dihedral angles strain energy

The dihedral angle, or torsion terms, have been presented in the discussion of the structural parameters of G-quadruplexes. This potential isessential for the description of the flexibility of polymer chains. Indeed, it allows the mutual rotation of chemical moieties [140].

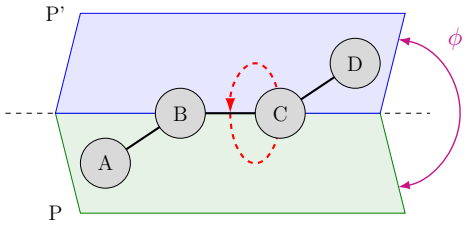


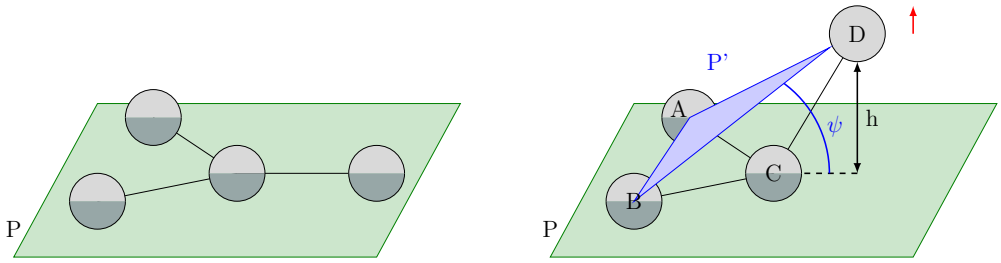
Figure 9.3: Schematic description of distortion of a dihedral angle, ϕ, between planes P and P', formed by the atoms ABC and BCD, respectively. The red dashed arrow symbolizes the rotation of atom D around the axis formed by atoms B and C. *Adapted from Chipot [143].*

The equation describing the potential energy for the rotation around a dihedral angle (E_{dihedral}) is shown below. It is a periodic function containing three parameters, known as i) the "height" of the torsion barrier ($V_n/2$), ii) the periodicity of the rotation – also called the multiplicity – which gives the number of minima that the function encounters during a complete rotation of 360° (n) and iii) the phase of the function (γ).

$$E_{\text{dihedral}} = \sum_n \frac{V_n}{2} \left[1 + \cos\left(n\phi - \gamma\right)\right] \tag{9.4}$$

9.3.4 Improper dihedral angle strain energy

Like the dihedral angle, the improper dihedral (or torsion) angle (ψ) is also established by four atoms, but arranged in a Y shape. In fact, it is essentially used to describe out of plane movement of atoms in coplanar systems, see Figure 9.4. Its usefulness becomes obvious when one wishes to force the coplanarity of the atoms constituting an aromatic ring, for example. Consequently, this potential is used to maintain an improper angle value as close as possible to 0 [133, 140]. Concretely, we consider four atoms ABCD chemically linked, where C is the central atom and D is the mobile atom. If the four atoms lie on the same plane P, then they are said to be coplanar (or in the *In-plane* configuration). But if the mobile atom D is outside the plane P, then the configuration of the four atoms is said to be *Out-of-plane*. The improper angle ψ is formed between the two planes: P and P', determined by atoms ABC and ABD, respectively.
MacKerell et al. [145] gives the following equation to calculate the potential of the improper dihedral angle, where k_ψ is the angular force constant at equilibrium.



(a) *In-plane* **(b)** *Out-of-plane*

Figure 9.4: Schematic description of the *In-plane* (a) and *Out of plane* (b) configurations and of the improper dihedral angle, ψ, between planes P and P', formed by the atoms ABC and ABD, respectively. *Adapted from Malliavin [144]*

$$E_{\text{Improper}} = k_{\psi} \left(\psi - \psi_0\right)^2 \tag{9.5}$$

9.3.5 Cross terms

Cross terms are potential energy functions that couple different energy terms. For example, the *stretch-bend* is a cross term that couples the elongation energy of a covalent bond with the angular strain energy. It is useful because the distance value of a covalent bond is often reduced during the opening of the valence angle. The equation describing the *stretch-bend* cross term is shown below, in which $k_{r\theta}$ is the equilibrium elongation and angular force constant. The use of cross terms, shown in Figure 9.5, allows to obtain a better description of the system dynamics. However their use heavily increases the computation time and furthermore requires often a consequent work of reparameterization of the force field [132, 133].

$$E_{\text{Bond-Angle}} = k_{r\theta} \left(r - r_0\right) \left(\theta - \theta_0\right) \tag{9.6}$$

9.4 Non-bonded interactions

Electrostatic and Van der Waals interactions concern both covalently bonded and non bonded atoms. These interactions can be described as shown in Figure 9.6:

- Intramolecular interactions, if they involve atoms of a molecule.
- Intermolecular interactions, if they involve atoms of different molecules.

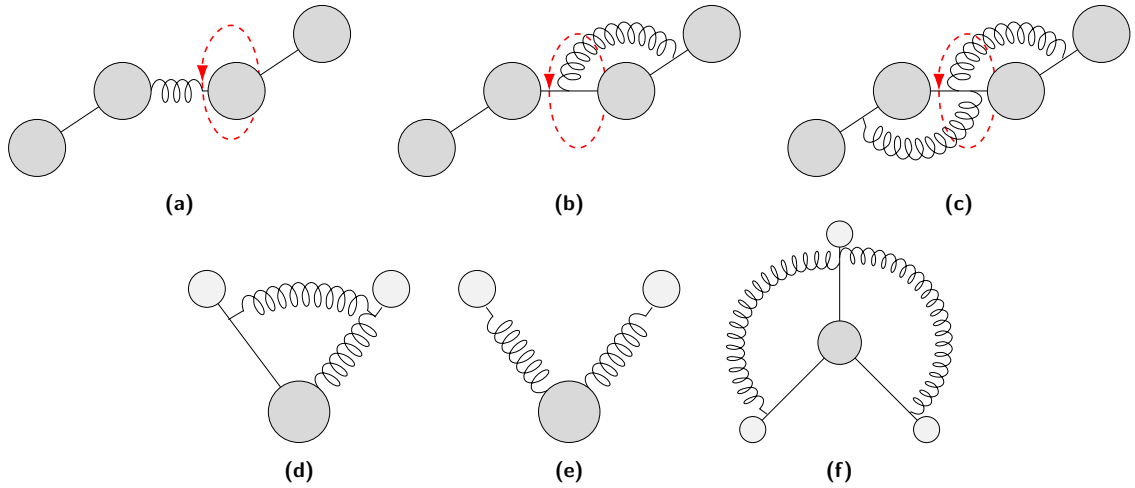


Figure 9.5: Examples of cross terms: stretch-torsion (a), bend-torsion (b), bend-bend-torsion (c), stretch-bend (d), stretch-stretch (e), bend-bend (f). *Modified from Chipot [143].*

The interactions between atoms connected by covalent bonds are described by the bond and angle strain energy terms described above. In this sense, covalently bonded atoms are *type 1-2* connected, atoms forming a bond angle are *type 1-3* connected and atoms forming a dihedral angle are *type 1-4* connected. Atoms with more than *type 1-4* connections can interact only through non-bonding interactions.

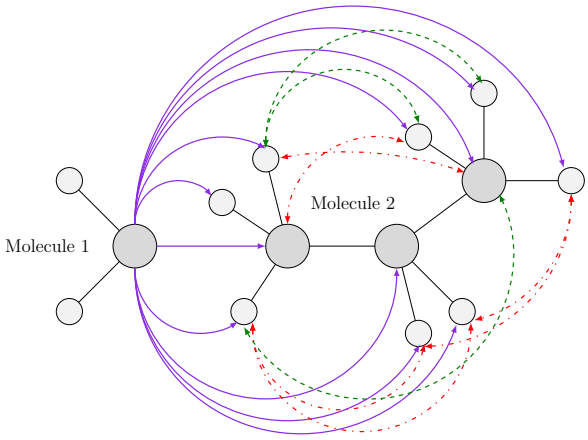


Figure 9.6: Possible interactions between non bonded atoms. The solid purple lines represent the intermolecular interactions; the green dashed lines represent the intramolecular interactions, within Molecule 2, involving atoms with more *type 1-4* connections; the red dash-dotted lines represent the intramolecular interactions, within the Molecule 2, involving atoms with *type 1-4* connections. *Modified from Chipot [143].*

The electrostatic potential energy function ($E_{\text{electrostatic}}$) involves atoms with partial point charges, or ions possessing a formal charge. The function is described by the classical Coulomb law. This involves the partial charges of each atom (q_i and q_j), the distance between them (r_{ij}), the dielectric permittivity of vacuum (ϵ_0) and the effective dielectric constant of the medium (ϵ_1) [133, 131].

$$E_{\text{electrostatic}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \tag{9.7}$$

Van der Waals interactions take into account the attraction-repulsion occurring between non-bonded atoms. The most common description of the Van der Waals potential energy function is the Lennard-Jones equation, also called *6-12 potential*, as follows:

$$E_{\text{vdW}} = \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] \tag{9.8}$$

Where, R^*_{ij} is the equilibrium distance between the two interacting atoms, and ε_{ij} corresponds to the minimum energy of the function (the depth of the potential well) [133, 131, 132]:

9.5 AMBER force field

The molecular dynamics simulations performed for the research presented in this thesis used the potential energy function ($E(r^N)$) of the AMBER (*Assisted Model Building and Energy Refinement*) suite of biomolecular simulation programs, whose detailed explanations have been provided by Case et al. [146] and Salomon-Ferrer, Case, and Walker [147]. Bonomi and Camilloni [140] has provided a mathematical description of this force field, i.e. of the whole potential energy function $E(r^N)$, in which each term has been discussed previously. A more thorough explanation and discussion of the force field used for proteins and nucleic acids can be found in articles written by Kollman et al. [148], Ponder and Case [149], Cornell et al. [150], Cheatham and Case [151], Maier et al. [152], and Tian et al. [153].

$$\begin{aligned} E(r^N) = & \sum_{\text{bonds}} k_r \left(r - r_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} \sum_{n} \frac{V_n}{2} \left[1 + \cos\left(n\phi - \gamma\right)\right] \\ & + \frac{1}{k^{1-4}_{\text{vdW}}} \sum_{\substack{i<j \\ \{i,j\}\in 1-4}} \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{k^{1-4}_{\text{Coulomb}}} \sum_{\substack{i<j \\ \{i,j\}\in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\ & + \sum_{\substack{i<j \\ \{i,j\}>1-4}} \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{\substack{i<j \\ \{i,j\}>1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \end{aligned} \tag{9.9}$$

In this equation, there are two terms describing the interaction between non-bonded atoms: one for those separated by three chemical bonds, the so-called 1-4 contribution, and the other for atoms separated by more than three chemical bonds. The 1-4 contributions of the Van der Waal and electrostatic interactions are weighted by $1/k_{\mathrm{vdW}}^{1-4}$ and by $1/k_{\mathrm{Coulomb}}^{1-4}$, respectively.

9.6 *Ab initio* parameterization of AMBER force field for a general molecule.

The underlying physical laws necessary for the mathematical representation of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact solution of these laws leads to equations much too complicated to be soluble.

Paul Dirac – 1929

The AMBER force field contains a large set of parameters for a large number of atoms and is able to describe the structure and dynamical evolution of several biological molecules. However, when considering a new molecule or unit, like a DNA lesions such as a strand breaks or modified nucleotide or small ligands, it is often necessary to introduce new parameters for describing the properties of atoms and the bonds between atoms non present in the force field. This is typically done by performing quantum chemical calculations on the new molecules.

9.6.1 Geometry optimization

The potential energy function $E(r^N)$, also called potential energy (hyper)surface, allows to link the energy of a molecule with its geometry, i.e. with the position of its N nuclei. And whatever the method used to describe this function, the objective of the geometry optimization is always to find minima on the potential energy surface function [154]. In fact, the minimum of the function corresponds to a stable molecular configuration, thus to an equilibrium arrangement of the nuclei. Cramer [133] points out that

geometric optimization is in fact a mathematical problem, and then asks the following question to explain his statement: « How does one find a minimum in an arbitrary function of many variables ? ». In fact, the algorithm acts on a geometrical parameter, also called degrees of freedom (for example, a link or an angle), while keeping the others constant. When changing a geometrical parameter causes an increase in energy, then this indicates that the algorithm is pushing in the wrong direction and needs to be changed. On the contrary, if the modification of the geometrical parameter induces a decrease in energy, then the algorithm is pushing in the right direction. The algorithm will continue in this way until the decrease in energy is no longer possible, and it will explore the next geometrical parameter, until all degrees of freedom are completely explored. This represents an optimization procedure. The objective of the algorithm is to perform a succession of optimization steps until a point is reached where the energy cannot decrease any more [133]. From a mathematical point of view, a potential energy minimum corresponds to a point on the potential energy surface for which the forces are zero and the first derivative of the energy, i.e. the gradient, is also zero. The force constants are second derivatives which can then be used for the harmonic terms in the force field. They are calculated at each step are arranged in a Hessian matrix, which allows to estimate the curvature of the potential energy surface function [154]. For example, the Gaussian [155] software may use an optimization tool based on the Berny algorithm [156, 157, 158].

9.6.2 Assignment of atomic charges

This strategy is specific for Amber force field. Charmm and other force fields have different strategies.

Restrained electrostatic potential (RESP) method allow the calculation of partial atomic charges based on the local electrostatic potential [133], obtained by a charge grid in Gaussian. The RESP method, used in the present thesis work, assigns net charges to the atoms of a molecule, whose sum provides the total net charge [159]. In the present work, it is used through the Antechamber tool of the Amber suite [146], for the parameterization of force fields of organic molecules [160].

9.6.3 Force constants

To calculate bond and bond angle force constants, the easiest way is to use the Parmchk2 tool implemented in the AmberTools suite [146]. Its purpose is to find the force field parameters: geometrical parameters (angles and distances) and the force constant. Parmchk2 is powerful, but it is not always possible to assign parameters. In this case there are several solutions. It is possible to search manually in the database file used by Parmchk2, for example in the GAFF2 file. Otherwise, it is necessary to do a bibliographic search to find the missing parameters. The last options are to make quantum calculations of vibration frequencies or to use IR spectroscopy on samples containing the new molecule under consideration. The force constant (k) obtained by IR spectroscopy can be deduced from the vibration frequency using Hooke's law [161], where U is the potential energy, ν_{vibr} is the vibration frequency, and m_1, m_2 are the masses of atoms 1 and 2:

$$U(r) = \frac{1}{2}k\,(r - r_{eq})^2 \tag{9.10}$$

$$\nu_{vibr} = \frac{1}{2\pi} \times \sqrt{\frac{k}{\frac{m_1 \times m_2}{m_1 + m_2}}} \tag{9.11}$$

On the other hand, software like Gaussian [155] provides force constants by diagonalizing the hessian matrix and the corresponding vibration frequencies, assign also the corresponding vibration normal modes. [133].

CHAPTER 10

Biased molecular dynamics

Chapter contents

Simulations calculated on the basis of equilibrium molecular dynamics provide information on interactions, or structural transformations, that take place on a microsecond time scale (μs). However, there are events which take place on a higher order time scale. In such a situation, equilibrium molecular dynamics is not properly describing the physical process we want to study. The solution to such a situation is to renounce to the explicit time description of the phenomenon and replace it by a description of thermodynamic properties such as free energy barriers and profiles. To do this a bias is applied to the molecular dynamic simulation to allow the system to overcome the free energy barriers leading to the desired event, knowing the biased potential applied will allow then to reconstruct the free energy profile. The application of this strategy is done by using accelerated sampling methods, which is also known under the name of biased molecular dynamics.

Accelerated sampling methods use one or more collective variables, also named reaction coordinate, to reduce the dimensionality of the free energy surface. Then on obtains free energy profile depending on a reduc ed number of collective variables, which provide the description of the desired physical phenomenon, this is also called the Potential

of Mean Force (PMF). Biased molecular dynamic simulation may be performed using classical potential or on quantum approaches and hybrid. In this chapter I will explain the fundaments of metadynamics, eABF and the meta-eABF methods, which has been used in this work at classical level, and umbrella sampling used at the QM/MM level for the study of reactive events.

10.1 Metadynamics

Every molecular system has an energy landscape that represents the different possible states of this system. Each possible state corresponds to a minimum in the potential energy function of the system. Metadynamics , introduced in 2002 by Laio and Parrinello [162], is used to accelerate the exploration of the energy landscape of a molecular system, To do so, this method accelerates the time scale of rare events by adding a bias potential around one or two collective variables. Concretely, a Gaussian repulsive potential is added to the energy of the system. In doing so, there is a repulsion of the system out of the initial energetic state, towards another one, by passing through a transition state [163]. To explain it differently, the method prevents the system from returning to states already explored by molecular dynamics. Interestingly minus the sum of the repulsive potential converges towards the final free energy potential.

The metadynamics can be represented in the following way: a ball is at the bottom of a large container and to make it go out, we empty sand bags one by one until the height of the added sand makes the ball go out of the container.

The accuracy of the metadynamics is dependent on two things:

- The definition of the height and width of the Gaussians, which determines the accuracy of the method.
- The choice of the collective variables, because a wrong choice can prevent the observation of the phenomenon or lead to a wrong estimation of the free energy[163, 164].

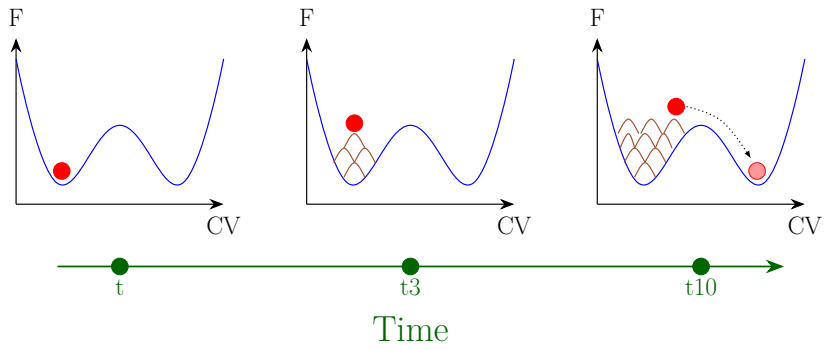


Figure 10.1: In a metadynamics calculation, the gaussians are added gradually over time. Little by little the system is repulsed and goes to another minima by passing through a transition state.

10.2 Difference in free energy between two states

Studies often seek to know the free energy between two states of a molecular system. In biology, this can be the energy required to make a structural change from a Y conformation of the protein to a Z conformation. For example, a protein channel can have two conformations: open or closed. The simplest method to achieve this is to perform a molecular dynamics to generate a set that contains a number of representative conformational configurations in the Y conformation (N_Y) and in the Z conformation (N_Z). Then difference of the free energy is calculated using the equation bellow [165]. This method has the advantage of being very simple. However, we can notice two important points limiting its implementation. First, it is necessary that the Y and Z conformations of the molecular system are not separated by a too large energy barrier. In fact, it is preferable to apply it to molecular systems whose $Y \rightarrow Z$ transition has a small energy barrier. Then, it is necessary that the frequency of presence of each event is sufficient in the set created. i.e. when the value of ΔF_{ZY} is low.

$$\Delta F_{ZY} = -k_B T ln \left[\frac{N_Z}{N_Y}\right] \tag{10.1}$$

10.3 Thermodynamic integration

When the system has a high energy barrier between the Y and Z conformations and/or when their frequency is low in the created set, then a thermodynamic integration can be performed. But, before giving the formula of thermodynamic integration, it is necessary to introduce the basis of the method. The Hamiltonian $H(p,r)$ of the system is also defined as a function of a coupling parameter (λ) between the two conformations of the molecular system: Y and Z. Consequently the Hamiltonian is written as $H(p,r,\lambda)$, where the parameter λ is chosen to correspond to the Hamiltonian of each conformation. This leads to :

- λ_Y corresponds to the Hamiltonian of conformation Y : $H(p,r,\lambda_Y) = H_Y(p,r)$
- λ_Z corresponds to the Hamiltonian of conformation Z : $H(p,r,\lambda_Z) = H_Z(p,r)$

The consequence of the expression of the Hamiltonian with respect to the parameter is that the derivative of of the free energy is also expressed as a function of this parameter. As a result, the free energy formula is written in the following equation [165]:

$$\frac{dF_\lambda}{d\lambda} = \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_i \tag{10.2}$$

In this equation, the system is forced to be confined according to a fixed value of lambda. As a result, it is necessary to calculate the average gradient of the Hamiltonian. This being said, we can explain the principle of thermodynamic integration. It is an empirical way to calculate the energy between state Y and state Z by integrating Equation 10.2 throughout the segment $[\lambda_Z, \lambda_Y]$ [165].

$$F(\lambda_Z) - F(\lambda_Y) = \int_{\lambda_Z}^{\lambda_Y} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_i d\lambda \tag{10.3}$$

The implementation of this method requires the use of a large number of calculation windows, in order to cover the whole energy profile of the molecular system. But this process is very costly in terms of computation time.

10.4 eABF and meta-eABF

To solve the computational time problem, other methods such as eABF and meta-eABF have been proposed. Their objective is to improve the formulation of the thermodynamic integration.

eABF method The Adaptive Biaising Force (ABF) method [166] directly biases the chosen collective variable. It is the origin of the extended Adaptive Biaising Force (eABF) method [167], yet the two methods differ. First, eABF relies on extending the molecular system q by a dummy collective variable λ, which fluctuates around the chosen collective variable ξ. This results in an extended molecular system (q, λ). λ has no physical character, is of mass m_λ, and changes progressively under the same temperature condition as the molecular system under study. Next, it is imperative to approximate the dummy collective variable λ to the chosen collective variable ξ. This step is done simply by coupling the two variables into a harmonic potential, namely : $\frac{k}{2}(\xi(q) - \lambda)^2$. After these explanations we can see that the application of the eABF method on a collective variable ξ, corresponds to the application of an ABF method on an extended collective variable ξ^{ext}. Here, it should be understood that $\xi^{ext}(q, \lambda) \equiv \lambda$ [167]. Finally, the eABF is a potential of mean force (PMF) method which has the advantage of being accurate and whose convergence can be calculated. However, it requires long computation times, as well as the establishment of computation windows.

meta-eABF method Fu et al. [168] propose a new framework to combine the advantages of metadynamics and eABF, in the strategy called meta-eABF. This method requires only a limited number of windows, in some cases just one, and provides accurate and converged results in usually shorter time then eABF. Concretely, the meta-eABF applies the metadynamics to sample the less important regions and applies the eABF on the more important regions.

10.5 Umbrella Sampling

Umbrella sampling is a method proposed by Torrie and Valleau [169] (see Figure 10.2). It varies a harmonic bias potential which is introduced on the coordinate of the path chosen to make the molecular system pass from state Y to state Z. Similar to other methods such as eABF, Umbrella sampling requires the use of several computational windows. But, here, it is very important to have a good overlap of the windows, i.e. between the distributions of the values of the chosen collective variable. Indeed, the sampling is done on the center of each window to provide a distribution set. The distribution sets of each window are biased by the introduced harmonic potential [170]. However, the energy profile of the $Y \rightarrow Z$ transition is found by applying the Weighted Histogram Analysis Method (WHAM) [171]. Concretely, WHAM combines all the windows of calculations and then removes the bias, which is known [172].

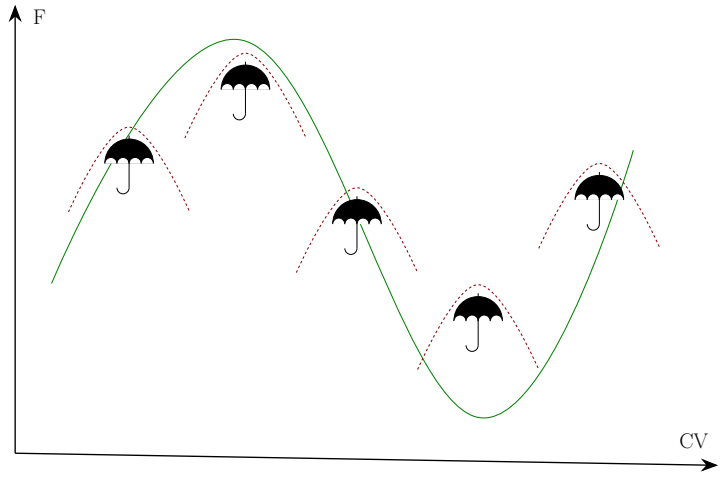


Figure 10.2: Schematic representation of the Umbrella sampling method. Each umbrella represents a calculation window.

CHAPTER 11

Methods for quantum chemical calculations

Chapter contents

Classical molecular dynamics is very adapted to the study of the conformational evolution of large molecular systems. However, this approach is useless when we are interested in enzymatic reactions, because it requires the study of chemical bond formation or breakage. The same is true for studying the spectroscopic properties of a molecule, because it is necessary to describe the electronic transition states. So when this type of study is considered, it is necessary to orient the calculations towards a quantum description of the molecular system, which necessarily involves solving the Schrödinger equation.

11.1 The Schrödinger equation

The classical approach describes atoms by considering only their nuclei while the electrons are treated implicitly via the force field. On the contrary, the quantum approach describes atoms by explicitly considering their nuclei and electrons. The quantum state of a system is a superposition of the configurations that this system can adopt. It is therefore impossible to describe precisely all the aspects of this system. The solution is to describe the system according to its wave function [173], which depends on the coordinates of nuclei and electrons (r). When we look for a stationary state independent of time $\Phi(\vec{r})$, the Schrödinger equation is written in the following terms [173]:

$$\hat{H}\Phi(\vec{r}) = E\Phi(\vec{r}) \tag{11.1}$$

In the time independent formalism, the Hamiltonian of a system composed of N electrons and M nuclei, contains the following terms: the electron kinetic energy ($\hat{T}_e$), the nuclear kinetic energy ($\hat{T}_n$), the interaction term between electrons and nuclei ($\hat{V}_{ee}$), the electrostatic repulsion between the electrons ($\hat{V}_{ne}$), the electrostatic repulsion between the nuclei ($\hat{V}_{nn}$). So the Hamiltonian equation may be written as:

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{ne} + \hat{V}_{nn} \tag{11.2}$$

The terms of the above equation can be decomposed according to: the mass of the nuclei (M), the atomic number of the nuclei (Z), the electrons (i and j), the distance between the electrons r_{ij} , the electron-nucleus distance (r_{iA}) and the distance between two nuclei (r_{AB}), which gives [133]:

$$\hat{H} = -\sum_{i=1}^{N}\frac{\nabla_i^2}{2} - \sum_{A=1}^{M}\frac{\nabla_A^2}{2M_A} - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>1}^{M}\frac{Z_A Z_B}{r_{AB}} \tag{11.3}$$

Solving the time-dependent Schrödinger equation gives the time-dependent wave function $\Psi(\vec{r}, t)$ of a system. In this case the equation is written [173]:

$$\hat{H}\Psi(\vec{r}, t) = i\hbar\frac{\partial}{\partial t}\Psi(\vec{r}, t) \tag{11.4}$$

It is possible to relate the time-independent steady state ($\Phi(\vec{r})$ with the time-dependent wave function ($\Psi(\vec{r},t)$), by the following formula [173]:

$$\hat{H}\Psi(\vec{r},t) = \Phi(\vec{r},t)^{\frac{-iEt}{\hbar}} \tag{11.5}$$

However, the Schrödinger equation cannot be solved analytically because of the correlated motions of particles [133]. That's why several forms of approximations have been proposed.

11.2 The BornOppenheimer Approximation

The motion of the nucleus of an atom is much slower than that of its electrons, so that the electronic relaxation can be considered as instantaneous compared to the motion of the atoms. Thus, it is possible to decouple the two motions. This leads to consider that the nucleus has a fixed position and that the electrons are in motion. The Born-Oppenheimer approximation is therefore to calculate the wave function based on the decoupling of the motion of the nucleus and that of the electrons. By reasoning in this way, the Schrödinger equation is written for a polyelectronic system. It uses several terms: the kinetic energies of the electrons, the attraction between nuclei and electrons, the repulsion between electrons and a constant for a given set of fixed nuclear coordinates (V_n). But in practice V_n is often ignored. Note that by solving the Scrhödinger equation for different nuclear arrangement one explicitly obtains a potential energy surface, whose importance has been described earlier. In conclusion, the Born-Oppenheimer approximation gives the following equation [133]. But, this approximation gives a multivariate N-electron function that remains impossible to solve for polyelectronic atoms

$$\left(-\sum_i \frac{\hbar^2}{2}\nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}}\nabla_i^2\right) \Psi\left(r_i, R_A\right) = E\Psi\left(r_i, R_A\right) \tag{11.6}$$

11.3 The Hartree-Fock approximation

Because electrons are fermions it's important to consider the spin-orbital asymmetry when exchanging two particles. So the polyelectron wave function (Ψ) is expressed as a Slater determinant [133, 131]:

$$\Psi\left(\xi_1, \xi_2, \cdots, \xi_i, \xi_j, \cdots, \xi_N\right) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\xi_1) & \chi_2(\xi_1) & \cdots & \chi_N(\xi_1) \\ \chi_1(\xi_2) & \chi_2(\xi_2) & \cdots & \chi_N(\xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\xi_N) & \chi_2(\xi_N) & \cdots & \chi_N(\xi_N) \end{bmatrix} \tag{11.7}$$

11.3.1 The Hartree-Fock approximation

The Hartree-Fock approximation proposes to solve the Schrödinger equation by describing separately the interaction of an electron with the nuclei and the mean local electron repulsion field. To do this, this approximation is based on a single determinant, each term of which represents an orbital. In fact, a "pseudo" monoelectronic operator replaces the search for the eigenvalues and eigenfunctions of the Hamiltonian operator. It appears that the Schrödinger polyelectronic equation can be rewritten as a set of equations based on monoelectronic Hamiltonians (h_i), the energy of the i orbital (ϵ_i) and the the set of spin orbitals i (spinorbit) (ξ_i) [133]:

$$h_i \chi_i(\xi_i) = \epsilon_i \chi_i(\xi_i) \tag{11.8}$$

Classically the Hartree-Fock equation is written according to the Fock operator ($\hat{F}$) and the monoelectronic space function ($\varphi_i(\vec{r})$) [174]:

$$\hat{F} \varphi_i(\vec{r}) = \epsilon_i \varphi_i(\vec{r}) \tag{11.9}$$

The Fock operator is the monoelectronic Hamiltonian and it is expressed as follows [174, 133]:

$$\hat{F}_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j\geq 1}^{N}\left(2J_j(r_i) - K_j(r_i)\right) \tag{11.10}$$

This writing considers two things. First, the electron Coulomb repulsion, which is expressed through a Coulomb operator (J_j). Secondly, the Pauli principle which is expressed under an exchange operator (K_j) [133, 132]. But this writing remains heavy. Also, it can be rewritten by considering the kernel operator. This is a monoelectronic operator which describes the evolution of an electron in the field of the nuclei. It is calculated by the sum of the kinetic energy and the coulombic attraction energy by the nuclei [131, 132]:

$$\hat{h}_i^c = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \tag{11.11}$$

The Fock operator is thus written as follows:

$$\hat{F}_i = \hat{h}_i^c + \sum_{i=1}^{N}\sum_{j\geq 1}^{N}\left(2J_j(r_i) - K_j(r_i)\right) \tag{11.12}$$

Finally, it must be pointed out that the Hartree-Fock approximation neglects the electronic correlation. This often leads to a very bad reproduction of the experimental data.

11.3.2 Concept of *atomic orbital basis*

An atomic orbital basis is a linear combination of functions centered on the nuclei of atoms. The functions are associated to the atomic orbitals and can be expressed in two ways. Either by the use of Slater function (Slater Type Orbital), or by the use of Gaussian function (Gaussian Type Orbital). [133, 132, 131].

Solving Schrödinger's equation means to reproduce exactly the molecular orbitals. To do this, it is necessary to use an infinite number of functions. That is to say that it would be necessary to use an infinite basic set. It is obvious that such a thing is technically impossible. This is why it is necessary to introduce approximations. Obviously, the larger the base set, i.e. with a large number of functions, the more precise the calculations will be, but the longer they will take. To answer the demand of this problem, several basis sets have been developed and then implemented in the code of quantum chemistry

software [133], like Gaussian [155]. As an example of base, we can quote 6-31G which uses a combination of six Gaussian for the internal electronic layers, and a combination of three Gaussian and a diffuse function for the valence electronic layers. While the base 6-31G**, more precise, adds 6 primitives of order 2 for the atoms other than hydrogen [133].

11.4 Density functional theory (DFT)

In 1998, the Nobel Prize in Chemistry was awarded to Walter Kohn *for his development of the density-functional theory* and to John A. Pople *for his development of computational methods in quantum chemistry* [175].

DFT is a method which consists in using the number of electrons per unit volume at the point in space with coordinate r, also named electronic density $n(r)$, of a N-electrons system to obtain its energy $E[n(r)]$. DFT method is based on two theorems established by Hohenberg and Kohn:

- In the ground state, the electron density $n_0(r)$ sets the external potential $V[n_0]$. Therefore, the electron density conditions the Hamiltonian of the system. Consequently, all observables of the system resulting from the electron density can be known: such as the wave function Ψ, the type of atom, or the electronic energy of the system E [133, 131].
- The minimization of the total system energy (variational principle) established for the wave function is transposable to the electron density. Thus if the spatial distribution of the electron density of the system is in the ground state, then the energy functional is minimal [133]:

$$\min_{n(r)\to n_0(r)} E[n(r)] = E[n_0(r)] = E_0 \tag{11.13}$$

In the DFT case, the Born-Oppenheimer approximation remains valid. So the Hamiltonian describing each system is written according to the same formula stated previously (Equation 11.3). For an N-electron system, the DFT problem is solved considering a fictious wavefunction representing non interacting particles but giving the exact electronic

density electronic density (Ψ), which determines the state of the system, and spin and space coordinates (τ_i) [131]:

$$n(r) = N \int |\Psi(\tau_1, \tau_2, \cdots, \tau_N))|^2 \, d\tau_r, d\tau_2, \cdots, d\tau_N \tag{11.14}$$

Then, it is important to specify the main characteristics of the electron density:

$$n(r \to \infty) = 0 \tag{11.15}$$

$$\int n(r) dr = N \tag{11.16}$$

The electronic energy of the system E is a density functional and is calculated in the following way, in the framework of DFT:

$$E = E[n_r] \tag{11.17}$$

In the ground state, the energy is a unique functional of n_0.

$$E = E[n_0] \tag{11.18}$$

The energy is expressed according to the kinetic energy ($\hat{T}[n(r)]$) and the electron/electron interaction energy ($\hat{V}_{ee}[n(r)]$), it is the term of repulsion between the electrons, and the electron/nucleus interaction energy ($\hat{V}_{ne}[n(r)]$) [131, 133]:

$$E[n_r] = \hat{T}[n(r)] + \hat{V}_{ee}[n(r)] + \hat{V}_{ne}[n(r)] \tag{11.19}$$

The equation is also written as [133]:

$$E[n_r] = \langle \Psi | \hat{T} + \hat{E}_{ee} | \Psi \rangle + \int dr \; v_{ext}(r) n(r) \tag{11.20}$$

For a system with N-electrons, kinetic energy ($\hat{T}[n(r)]$) and the electron/electron interaction energy ($\hat{V}_{ee}[n(r)]$ or $\hat{E}_{ee}$) are written as [133]:

$$\hat{T} = -\sum_{i=1}^{N} \nabla^2 \tag{11.21}$$

$$\hat{E}_{ee} = -\sum_{i=1}^{N}\sum_{j<1}^{N}\frac{1}{r_{ij}} \tag{11.22}$$

The addition of the two terms T and H gives rise to the universal Hohenberg-Kohn functional ($F_{HK}[n(r)]$). It can be written simply as follows [133]:

$$F[n(r)] = \langle\Psi[n(r)]|\hat{T} + \hat{E}_{ee}|\Psi\rangle \tag{11.23}$$

Which means that the energy is written:

$$E[n_r] = F[n(r)] + \int dr\ v_{ext}(r)n(r) \tag{11.24}$$

However, there is a complication: $F[n(r)]$ is unknown. The result is that $E[n(r)]$ cannot be calculated exactly. Fortunately, it is possible to solve this difficulty thanks to the second Hohenberg-Kohn theorem, which was mentioned above. Here, we simply recall that the ground state energy (E_0) is obtained for the electronic density when the system is at the ground state of ($n_0(r)$).

11.4.1 The Kohn-Sham equations

Kohn and Sham [176] calculate the energy functional of the system using an expression close to the Hartree-Fock, and making use of the fictious wavefunction following the equations (Equation 11.10). The only difference is that the exchange potential, in Hartree-Fock, is replaced by an exchange-correlation potential (V_{xc}) in the Kohn-Sham method [133].

$$\hat{F}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \int\frac{n(r')}{|r-r'|}dr' + V_{xc} \tag{11.25}$$

The Kohn-Sham operator ($\hat{F}_i^{KS}$) replaces the Fock operator ($\hat{F}$) in the Hartreee-Fock equationEquation 11.9, which gives:

$$\hat{F}_i^{KS}\varphi_i(\vec{r}) = \epsilon_i\varphi_i(\vec{r}) \tag{11.26}$$

So, in the Kohn-Sham theory the electronic energy of the fundamental state of the system is written [133]:

$$E[n(r)] = -\sum_{i}^{N} \frac{1}{2}\nabla^2 - \sum_{i}^{N}\sum_{k}^{M} \frac{Z_A}{r_{iA}} + \sum_{i}^{N} \frac{1}{2} \int \frac{n(r')}{|r - r'|} dr' + E_{xc}[n(r)] \tag{11.27}$$

Now it is necessary to explain that the exchange-correlation potential (V_{xc}) is the functional derivative of the exchange-correlation energy (E_{xc}), so it is written [133]:

$$V_{xc}[n(r)] = \frac{\delta E_{xc}[n(r)]}{\delta[n(r)]} \tag{11.28}$$

From a formal point of view, the Kohn-Sham equation is exact. However, $E_{xc}[n(r)]$ contains a kinetic contribution. But this poses a consequent problem: it cannot be known. Consequently this term must be approximated, this is the role of the functionals which have been developed for this purpose [133]. There are different families of functionals, some of them are detailed in the next subsection.

11.4.2 The three different families of functions

Local density approximation (LDA). It is the simplest functional. Because it uses the perfect gas model. This requires the assumption of a uniform electron density. The disadvantage of its use is that such an assumption misrepresents molecular and most atomic systems. In fact, it does not adequately reproduce the weak interactions: van der Walls and hydrogen bonds. Therefore, the dissociation energies are overestimated, while the bond lengths are underestimated [133].

Generalized Gradient Approximation (GGA). This type of functionals will introduce information on the $\nabla n(r)$ charge density gradient. This results in a better estimation of the binding energies, compared to the LDA [133].

Hybrid functionals. This type of exchange-correlation functional allows to correct the self interaction error generated in DFT. The introduction of percentage of the Hartree-Fock exact exchange eliminates the repulsion of an electron by itself. We can cite the example of B3LYP which is one of the most used hybrid functional[133].

11.5 Quantum mechanics coupled to molecular mechanics (QM/MM)

In the previous sections, the advantages and limitations of the classical and quantum approaches have been seen. The first one allows to simulate the structural behavior of large systems, the second one is able to simulate the formation and breaking of chemical bonds of smaller systems. But what to do when we want to study the reaction of an enzyme, for example? The solution is to use a hybrid method, named Quantum Mechanics/Molecular Mechanics (QM/MM), which aims at coupling the quantum with the classical approach. To do this, the method separates the system into two regions (Figure 11.1). One corresponding to a small number of atoms on which a quantum approach is performed (QM) to simulate the chemical reaction, while The other one includes all the other atoms on top of which a classical approach is used (MM), because they do not participate to the chemical reaction [177]. So, the advantage of QM/MM is the separation of the MM and QM parts of the system. In other words, the classical approach is performed only on the MM region, keeping the accuracy of the quantum approach on the QM region, which is the site of interest of the chemical reaction.

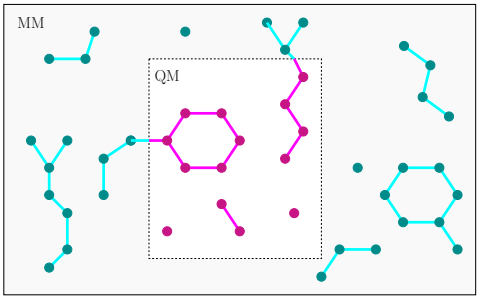


Figure 11.1: Representative diagram of a QM/MM coupling scheme.

In QM/MM the system is characterized by an effective Hamiltonian, which consists of three terms: QM Hamiltonian, MM Hamiltonian and coupling QM/MM Hamiltonian. Depending on the case, the QM/MM Hamiltonian can be added or subtracted to the other two. So the possible equations to obtain the effective Hamiltonian are [133]:

$$\hat{H}_{eff} = \hat{H}_{MM} + \hat{H}_{QM} + \hat{H}_{QM/MM} \tag{11.29}$$

$$\hat{H}_{eff} = \hat{H}_{MM} + \hat{H}_{QM} - \hat{H}_{QM/MM} \tag{11.30}$$

However, QM/MM presents a problem. Since the two parts of the system are treated at different level of theory, it becomes difficult to link the two parts together. This remark is more understandable if we consider a molecule where some atoms are located in the MM part, while the others are located in the QM part. For example, in the case of a nucleotide belonging to a DNA, it is possible to put the base in the QM part and the backbone in the MM part. By doing so, we realize a cut of the molecule at the level of the C — N bond between the sugar and the base. So two atoms involved in the same chemical bond are in the QM and tin the MM region, respectively. Thus, the problem comes from the description of the system at the border between the QM and the MM part [133].

To answer this problem, the simplest solution is to add a monovalent atom, usually hydrogen, to the center of the bond located at the boundary [133] to avoid unpaired electrons or unphysical charge distributions. This is a procedure proposed by M. Karplus. So, if the A — B bond is at the boundary, the application of this solution results in having an A — H in the MM part and an H — B bond in the QM part. Generally, it is better to add the atom in a bond where the A and B atoms are of the same kind, for example a C — C bond.

The use of QM/MM methods was first proposed in 1976 by Warshel and Levitt [177] and was the subject of a Nobel Prize awarded in 2013 to Martin Karplus, Michael Levitt and Arieh Warshel *for the development of multiscale models for complex chemical systems* [178].

4

Modeling and simulating G-quadruplexes

CHAPTER 12

in vitro model validation of G-quadruplexes structure

The study of nucleic acids often requires to know their three-dimensional structure. Several experimental techniques allow to access it, among the other NMR, X-ray crystallography, or crioelecrtonmicroscopy. However, not all structures are known, either because no study has been conducted, or because of experimental constraints. Computational approaches are able to predict the three-dimensional structures. Here, we can quote AlphaFold [179] which uses Deep Learning to predict the structure of proteins from their sequence. Unfortunately there is no software as powerful as this one for G-quadruplexes, and there are comparatively fewer structure solved. For example, a search in the NDB database [180], specifying only *Quadruple Helix* in the *Nucleic Acid Structural Conformation* section, returns 365 DNA and 38 RNA G-quadruplexes for a total of 403 structures. So if the structure of a G-quadruplex is not known, it becomes necessary to use a strategy to determine it. This is what is presented in the following article. More precisely, the research focuses on an *in silico* approach explaining the construction and validation of a model of G-quadruplex structure formed by the RG-1 RNA sequence of the SARS-CoV-2 genome. The model is first constructed using molecular homology and validated using a combination of classical molecular dynamics and QM/MM modeling. The validation strategy is based on the correspondence between the circular dichroism spectrum obtained experimentally by Zhao et al. [58] and the one that has been calculated on top of the structures obtained by our molecular dynamic simulations thanks to QM/MM TD-DFT calculations. In this approach the whole G-quadruplex core has been considered in the QM partition, other ways to calculate a theoretical circular dichroism spectrum of a macromolecular system exist based on Fragment Diabatization-based Excitonic (FrDEx) or Frenkel excitonic model [181].

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses

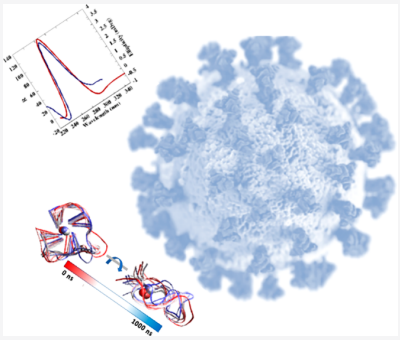
Tom Miclot, Cécilia Hognon, Emmanuelle Bignon, Alessio Terenzi, Marco Marazzi, Giampaolo Barone,* and Antonio Monari*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 10277−10283

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Guanine quadruplex (G4) structures in the viral genome have a key role in modulating viruses' biological activity. While several DNA G4 structures have been experimentally resolved, RNA G4s are definitely less explored. We report the first calculated G4 structure of the RG-1 RNA sequence of SARS-CoV-2 genome, obtained by using a multiscale approach combining quantum and classical molecular modeling and corroborated by the excellent agreement between the corresponding calculated and experimental circular dichroism spectra. We prove the stability of the RG-1 G4 arrangement as well as its interaction with G4 ligands potentially inhibiting viral protein translation.



At the end of 2019, a sudden outbreak of Severe Acute Respiratory Syndrome (SARS) developed in mainland China[1] and further spread worldwide, obliging the World Health Organization (WHO) to declare the emergence of a pandemic in March 2020. The syndrome is caused by a novel coronavirus, SARS-CoV-2, and has been styled COVID-19.[2,3] Despite the relatively low mortality, SARS-CoV-2 is highly contagious, and COVID-19 can evolve into severe forms necessitating critical care. Hence, COVID-19 is causing a considerable strain on healthcare systems, requiring unprecedented large-scale social distancing and containment measures, including full lock-downs. Even though vaccines have been promptly developed and released,[4] including the emergent mRNA (mRNA) technology,[5] COVID-19 is still raging worldwide as of summer 2021, pushed by the emergence of more contagious viral strains, such as the Δ variant, which has become notably dominant in Europe and the United States. The harmful effects of COVID-19 are also aggravated by the fact that no clearly efficient and safe antiviral agent has been proposed for large-scale use. Indeed, while some nucleoside analogues, including Remdesivir, have shown high antiviral efficiency *in vitro* and *in vivo*, related side-effects strongly hamper their diffusion.[6] In parallel to the mobilization of the medicinal chemistry community,[7] several structural biology[8] as well as molecular modeling and simulation[9] groups have produced an unprecedented effort, which has allowed the resolution and characterization of the main SARS-CoV-2 structural and nonstructural proteins.

The rather complex organization of the viral genome, also in the case of RNA viruses, has been recently highlighted and related to their biological functions. In particular, the presence of guanine-quadruplexes (G4) arrangements has been spotlighted.[10]

G4 arrangements are common in guanine-rich regions and are achieved by a superposition of planes composed of four guanines (tetrads) which are stabilized by the formation of Hoogsten-type hydrogen bond and by the interaction with cations placed in the central canal.[11,12] G4s have been recently associated with important roles,[13−15] such as the protection of telomeric regions in DNA[16] and the control of gene expression, and have also been highlighted in viruses.[17] The presence of quadruplexes' folding may preserve the viral genetic material, avoiding its recognition by the immune system. On the other hand, it has been shown that an overstabilization of G4s may inhibit the translation of viral proteins by the cellular apparatus. In the case of SARS-type coronaviruses, it has also been shown that the highly conserved SARS unique domain (SUD), used to sequestrate pro-apoptotic cellular mRNA sequences, is maintained in its active dimeric form by the interaction with G4 RNA sequences.[18−20] In fact, while SARS-CoV-2 genome is to some extent less prone to arrange in quadruplexes, compared to other viruses such as Zika,[21] four putative G4 sequences have been recently

Received: September 17, 2021
Accepted: October 14, 2021
Published: October 15, 2021

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

The Journal of Physical Chemistry Letters pubs.acs.org/JPCL Letter

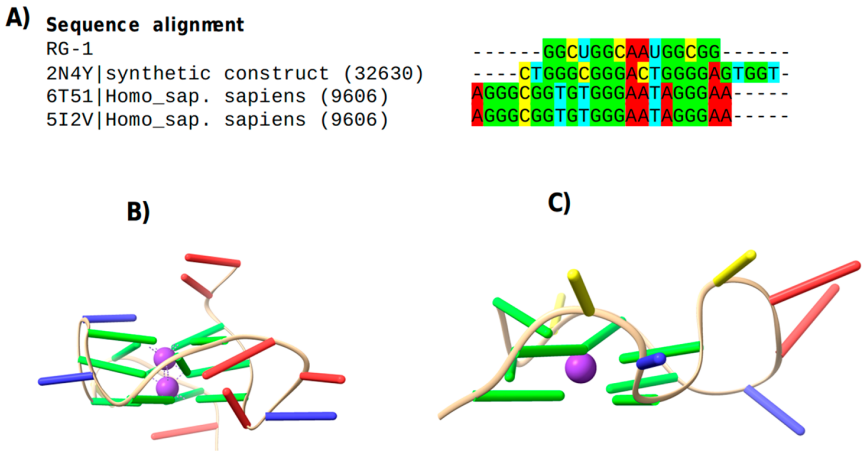


Figure 1. (A) Alignment of RG-1 sequence with a synthetic construct from HIV-1 genome (2N4Y) and two *Homo sapiens* DNA G4s sequences (6T51 and 5I2 V). (B) Side view of the structure of the G4 DNA from PDB 6T51 and (C) of the reconstructed RG-1. The stabilizing K^+ ions, which are included in the simulation, are represented in purple.

evidenced by Zhao et al.[22] In particular the so-called RG-1 sequence, located in the nucleocapsid (N) protein coding region, has been characterized using electronic circular dichroism (ECD). The presence of G4s in infected living cells has also been confirmed, and their stabilization by ligands can induce the downregulation of their expression, impairing the maturation and infectivity of viral proteins, hence paving the way to appealing therapeutic strategies.

Despite the importance of RNA G4 in the biological cycle of viruses like SARS-CoV-2, their structural characterization usually remains elusive. This is also underlined by the relative scarce number of RNA G4 structures that have been resolved, especially compared to their DNA counterparts. At the same time, the high ECD sensitivity to secondary structure rearrangements allows achieving a molecular resolution giving access to all the subtle structural factors that may be crucial in driving the possible interactions with external ligands. In this Letter, we study the RG-1 sequence through a combination of multiple sequence alignment, homology modeling, classical molecular dynamics (MD), and hybrid quantum mechanics/molecular mechanics (QM/MM), in order to disentangle all the structural factors associated with the G4 conformations, and by comparing the time-dependent density functional theory (TD-DFT) simulated ECD spectrum with the experimental one by Zhao et al.[23] The full computational strategy of our multiscale approach can be found in the Supporting Information.

The structure of RG-1 was predicted from a multiple sequence alignment with three DNA G4 sequences corresponding to experimental structures (PDB codes 2N4Y,[24] 6T51,[25] 5I2V).[26] The RG-1 structural model exhibits a parallel G4 composed of two superposed tetrads (see Figure 1). Interestingly, we can also evidence that, in addition to the G4 core, a rather large nonstructured loop, composed of adenines and cytosines, is also present as a linker to the guanines involved in the tetrad. The same loop is also present on the human DNA G4 used as template (PDB 6T51) (see Figure 1B). Three independent 1 μs MD trajectories of the RG-1 G4 model solvated in a periodic box of TIP3P water[27] have been run using the amber f99 force field including the parmbsc0[28] and the χOL3[29] corrections to take into account the specific RNA features. In addition simulations performed modeling the nucleic acid with the bsc1 corrections,[28,30] i.e., with force field specifically designed for DNA, have also been performed. All the simulations have been carried out with the NAMD code.[31] For sake of clarity, the results of all the replicas are collected in Table 1, in the form of average and standard deviation of the most crucial structural parameters. The time series of the same parameters are also presented in Figure 2 for the first replica while the other two, which exhibit identical trends, can be found in the Supporting Information.

From our MD simulations, the RG-1 tetrad core is extremely stable and experiences only slight vibrational deformation (see

Table 1. Average and Standard Deviation of the Main Structural Parameters for the RG-1 Simulations in the Three Replicas[a]

	replica 1	replica 2	replica 3
RMSD nucleic (Å)	4.66 ± 1.00	5.46 ± 1.07	5.79 ± 1.07
RMSD tetrads (Å)	1.63 ± 0.33	1.71 ± 0.24	1.81 ± 0.33
tetrad distance (Å)	3.43 ± 0.086	3.46 ± 0.08	3.46 ± 0.10
Ω (deg)	31.16 ± 2.42	31.30 ± 2.53	31.13 ± 2.44
		α Tetrad	
$\theta_{G1\text{-}G15}$ (deg)	89.95 ± 4.29	92.70 ± 4.89	91.36 ± 4.43
$\theta_{G11\text{-}G14}$ (deg)	88.96 ± 4.32	89.03 ± 4.80	90.64 ± 5.17
$\theta_{G5\text{-}G11}$ (deg)	89.14 ± 4.44	89.45 ± 4.49	89.07 ± 4.39
$\theta_{G5\text{-}G14}$ (deg)	92.80 ± 4.09	89.88 ± 4.24	90.21 ± 4.62
$\theta'_{G1\text{-}G14}$ (deg)	170.15 ± 5.81	170.26 ± 5.71	169.73 ± 6.76
$\theta'_{G1\text{-}G11}$ (deg)	167.97 ± 6.82	165.98 ± 7.76	165.88 ± 8.59
		β Tetrad	
$\theta_{G2\text{-}G6}$ (deg)	88.50 ± 3.98	90.94 ± 4.67	89.48 ± 4.24
$\theta_{G12\text{-}G15}$ (deg)	89.18 ± 4.61	90.39 ± 4.65	89.52 ± 4.49
$\theta_{G6\text{-}G12}$ (deg)	90.44 ± 4.21	91.19 ± 4.48	90.33 ± 4.10
$\theta_{G6\text{-}G15}$ (deg)	94.31 ± 4.51	92.29 ± 4.46	93.71 ± 7.89
$\theta'_{G2\text{-}G15}$ (deg)	165.90 ± 9.61	163.52 ± 9.97	163.89 ± 12.94
$\theta'_{G2\text{-}G12}$ (deg)	163.14 ± 9.62	159.83 ± 10.22	164.21 ± 9.42

[a]The values of the θ and θ' angles are reported for each of the couple of guanines constituting the tetrads. The distance between the tetrads is considered as the distance between the center of mass of each quartet.

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

The Journal of Physical Chemistry Letters pubs.acs.org/JPCL Letter

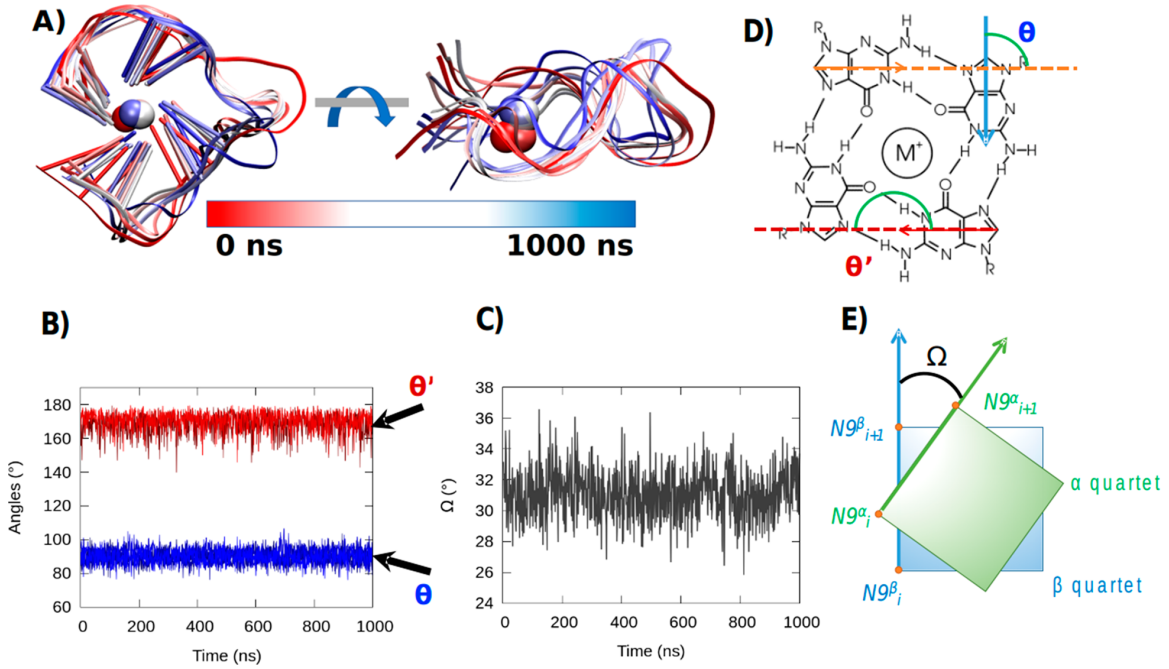


Figure 2. (A) Representative snapshots of the RG-1 RNA G4 extracted along the MD trajectory for the first replica. The color code represents the time evolution. Time evolution of the θ and θ' angles for the first, α, tetrad (B) and of the twist angle Ω (C). Schematic depiction of θ and θ' angles indicating the arrangement of guanines on the tetrad (D) and of the twist angle Ω (E). The results for the other replica and for the bsc1 force field are found in the Supporting Information and in Table 1.

Figure 2A). As expected, the connecting loop is much more flexible and accesses a larger conformational space. Interestingly, the fluctuations of this loop also lead to some metastable states in which adenines and cytosines can be π-stacked to the guanines in the tetrad.

Core rigidity and loop flexibility can also be inferred from the analysis of the root mean square deviation (RMSD) which is larger for the whole nucleic acid system than for the tetrad only. While the RG-1 sequence presents a moderate RMSD peak at around 8.0 Å, this value is mainly due to the contribution of the peripheral loop, which amounts to 5.8 Å. In contrast, the RMSD of the nucleobases in the tetrads barely reaches 1.8 Å, highlighting the negligible deviation of the tetrad core from the original starting conformation. As usual in G4 arrangements, the rigidity of the central core is due to the involvement of the guanines in the Hoogsteen hydrogen bonding network and to the favorable interaction with the metal cations present in the central channel. Of note, the same rigidity of the central core, combined with the flexibility of the peripheral loops, can also be observed from the 2-dimensional (2D) RMSD map reported in the Supporting Information.

The stability of G4 arrangement is reflected in the almost ideal values assumed by the θ and θ' angles defining the disposition of the guanines in the tetrad (see Figure 2D and Table 1). θ peaks at around 90°, and θ' is centered around 160° in both tetrads (see Figure 2-B, Table 1, and the Supporting Information). Furthermore, the standard deviations are relatively small, not exceeding 5°, and the values are coherently reproduced for all the replicas (Table 1). Such a situation is indicative of the formation of a stable and persistent Hoogsteen hydrogen bonding network, hence confirming the propensity of the RG-1 RNA sequence to assume a G4 parallel conformation. As a more global descriptor, the stability and rigidity of the G4 core is also reflected by the time evolution of the Ω angle describing the twisting between the two tetrads (Figure 2C,E), which is centered around the ideal value of 31° typical of a parallel arrangement, while the standard deviation is close to 2°. Importantly, once again the distribution of the parameter is coherently reproduced in all the replicas (Table 1 and the Supporting Information).

Behaviors similar to the ones reported in Figure 2 and Table 1 are observed when using the DNA-based bsc1 force field (see the Supporting Information), providing further proof of the robustness of our approach. Furthermore, by performing clustering of the trajectories of the independent replicas, we also confirm that the conformational space explored by our system is strongly overlapping.

Hence, our MD trajectories are coherent with the experimental data in confirming the possible folding of the RG-1 sequence of the SARS-CoV-2 genome into a parallel G4 conformation. In their original work, Zhao et al.[23] used ECD spectroscopic signatures to confirm the structuration of the RNA sequence into a G4 arrangement. To allow a better mapping between molecular modeling and experimental results we have simulated the ECD spectrum using a hybrid QM/MM approach on top of snapshots extracted from the MD trajectory. The simulation of ECD spectra from snapshots obtained from a classical sampling of the chromophore phase-space has been successfully used by our group for related systems;[32−34] however, in the previous contribution the rotatory strength and the excitation energies of the multi-chromophoric aggregates have been obtained considering an excitonic coupling Hamiltonian. Although powerful, the excitonic model is inherently semiempirical, especially in its dipole approximation, hence leading to a suboptimal

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

The Journal of Physical Chemistry Letters pubs.acs.org/JPCL Letter

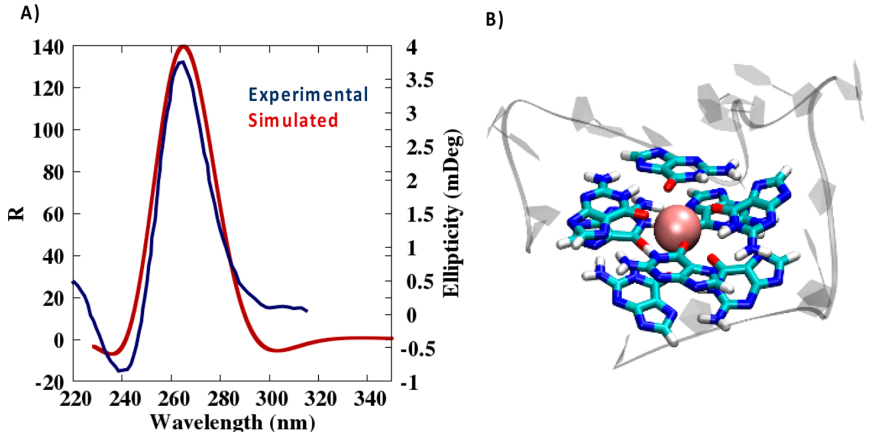


Figure 3. (A) Experimental, from Zhao et al.,[23] and simulated ECD spectrum of the RG-1 RNA sequence. Note that the simulated spectrum has been homogeneously shifted by 40 nm. The simulated spectrum is obtained by QM/MM at the TD-DFT level with M06-2X functional and the 6-31G(d) basis, calculating 60 excited states. (B) The chosen QM partition is highlighted in ball and stick representation, while the G4 backbone and dangling bases are shown in cartoon from.

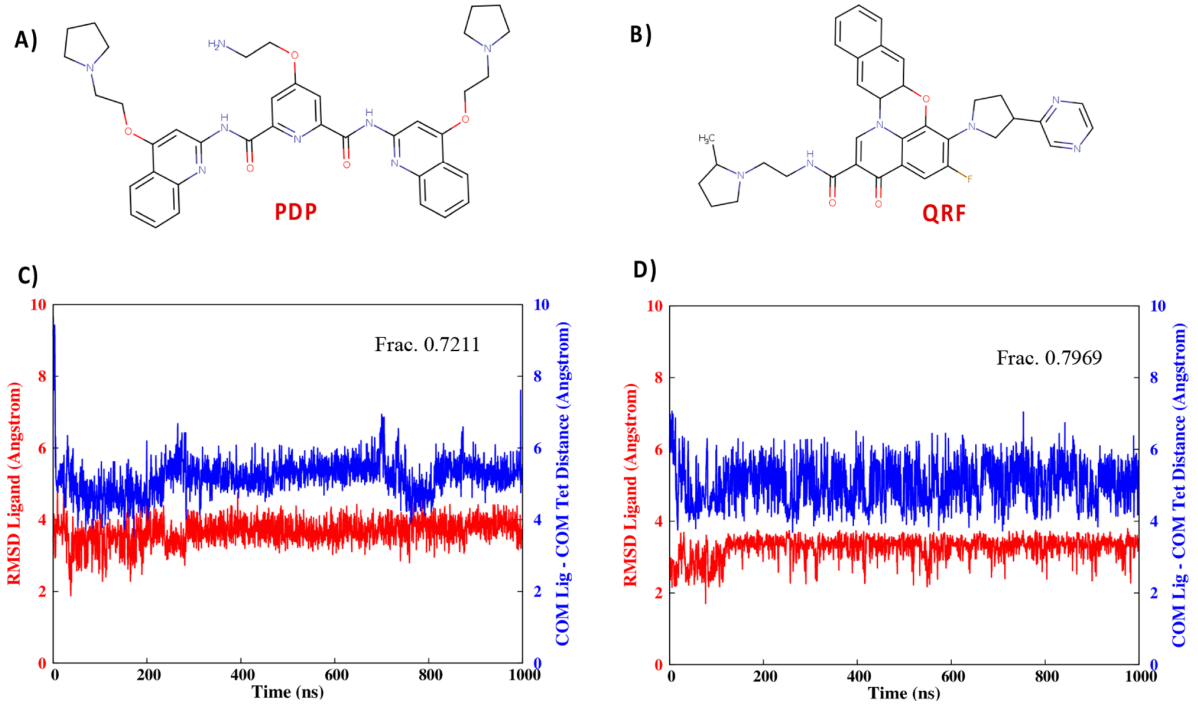


Figure 4. Chemical representation of quarfloxin, QRF (A), and pyridostatin, PDP (B), together with the most important conformation obtained from clustering of the ligand bound to G4, the time evolution of the RMSD, and the time evolution of the distance between the center of mass of the ligands and of the tetrads (C and D for QRF and PDP, respectively). Note that the RMSD is calculated from the starting frame.

description of relatively strongly coupled multichromophores. Furthermore, some dependency on the partitioning of the system can still be observed in the final results. For these reasons, we decide to bypass this problem by using a fully *ab initio* approach and including all the stacked nucleobases in the QM partition. The QM/MM excited-state calculations have been performed using the Orca/Amber interface, and the eight nucleobases forming the G4 core have been included in the QM partition to consider the coupling between the π-stacked chromophores (Figure 3B). Although we have applied a shift to the calculated absorption wavelengths, the results, in terms of band shape, are in good agreement with the experimental spectrum published by Zhao et al.[23] (see Figure 3). The experimental spectrum presents a first large and rather intense positive peak centered around 270 nm followed by a negative peak at 240 nm, whose intensity is much smaller, the ratio being about 1:4. Such a pattern is common in the case of nucleic acid aggregates and can be seen as a typical spectroscopic feature of parallel G4 arrangement. The results of our QM/MM simulation produced an ECD spectrum with a similar pattern, but with blue-shifted signals, due to the use of a reduced basis set.

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

Although a larger basis set could reproduce with more precision the absorption energy, in the context of the present work we considered of more prominence the correct description of the global shape of the ECD signal, thus confirming the G4 conformation, at the same time correctly describing the electronic nature of the spectral signature.

The choice of the basis set was imposed not only by the computational overload due to the extended QM partition but also by the necessity to avoid QM wave function over-polarization due to the interaction with the MM point charges. To palliate the basis set incompleteness, we apply a global shift of 45 nm to the simulated ECD signal to allow the straightforward comparison with the experimental counterpart (see Figure 3). As for the choice of the exchange−correlation functional in Figure S4 we report the simulated ECD spectra, obtained on top of the *bsc*1-based trajectory with hybrid, long-range corrected, and meta-hybrid functionals, illustrating the superiority of M06-2X over the other choices. The similarities between the simulated and experimental signals in Figure 3 are self-evident, providing a clear picture in terms of the relative position of the two main peaks, their intensity, and the global band shape: a broad positive band is followed by a less intense, and slightly sharper, negative signal. However, the higher-energy negative band has a reduced intensity probably due to the necessity of calculating a larger number of excited states. The globally good agreement between the calculated and experimental spectra, despite the applied energy shift, supports the folding of the RG-1 sequence in a parallel G4 conformation.

Having shown the stability and the persistence of the G4 arrangements, the question arises whether some ligands could interact with the RG-1 RNA sequence and possibly over-stabilize it to induce an effective inhibition of the translation of the viral genome and impair the viral cycle. In this context, we have examined the interaction of two ligands with the RG-1 sequence, namely, quarfloxin (QRF) and pyridostatin (PDP). QRF (also called CX-3543) is a well-known G4-binding ligand, believed to target RNA G4s and evaluated in phase II clinical trials for human cancer therapy.[35] PDP is also known for its capacity to modulate the expression of G4 containing genes, providing antiproliferative effects.[36] Interestingly, PDP was reported by Zhao et al.[23] as a ligand capable of selectively recognizing RG-1 and increase its melting temperature. The initial complexes between RG-1 and the two ligands have been obtained by docking the drugs onto the parallel G4 structure. The stability of each significant pose has been further confirmed running two independent 1 μs MD trajectories. Representative structures of the main binding modes are reported in Figure 4, and the results for a slightly different initial pose, in which the ligand interacts with the tetrad, are also collected in the Supporting Information.

Both the docking and the MD simulations agree in previewing the formation of stable complexes between the RNA G4 and the small ligands, the persistence and stability of which can also be appreciated by the extended plateau observed on the RMSD time-series after 400 ns (Figure 4B,D). More importantly, it can be observed that the binding takes place via π-stacking of the ligand on top of the tetrad plans, and it is mostly driven by dispersion and hydrophobic interactions. This result can be rationalized by the presence of conjugated moieties on the ligands and their globally planar structure. Furthermore, such interaction modes, while providing an enhanced stabilization of the G4 arrangement due to the increase of the attractive interactions, induce only moderate or negligible structural perturbation as can also be observed from the angles between the guanines and the twist reported in the Supporting Information for the different interaction modes and ligands. The structural stability of the G4 arrangement upon binding with ligands is also consistent with the experimental results that showed only slight differences in the ECD spectra upon interaction of RG-1 with PDP. In addition, the π-stacking interaction, also due to the largely aromatic nature of both ligands, is susceptible to take place without any important energetic barrier, or steric hindrance, hence greatly facilitating the recruitment of the drug. While incontestably pointing toward the formation of stable complexes, the results of the equilibrium MD simulations alone cannot underlie any significant or qualitative difference between QRF and PDP binding, hence suggesting that both can potentially be seen as valuable ligands to stabilizing G4 arrangements and more particularly the parallel conformation of the RG-1 RNA sequence. The stability of the complex and the presence of a dominant structure can also be evidenced by the clustering and the 2DRMSD reported in the Supporting Information. Indeed, as shown in Figure 4B,D we may see that the distance between the center of mass of the ligand and of the G4 tetrad is remarkably stable all along the trajectory. Notably, the dominant conformation obtained by clustering and corresponding to a bound complex is also accounting for about 80% for both drug candidates. However, the dominant interactions exhibited by the two ligands are rather unspecific, mainly driven by π-stacking, and should be taken as a proof of concept of the possible stabilization of the G4 structure. From a medicinal and drug-design point of view, it seems reasonable to infer that common, and known, G4-ligands could potentially affect SARS-CoV-2 replication. Furthermore, the increase of the π-conjugation and the presence of the fused aromatic ring, as well as a planar structure, appears as the most important characteristics for a potentially interesting lead compound.

We have unravelled the structural behavior of a putative G4 RNA sequence present in the genome of SARS-CoV-2 using multiscale molecular modeling approaches. The combination of the multiple sequence alignment, the microsecond-scale sampling of the conformational space, and computational spectroscopy support the fact that the RG-1 sequence may adopt a stable parallel G4 conformation composed of two rigid tetrads and a flexible peripheral loop. Our results point to the fact that the rigid core is reliably sampled, in the limits of the force field accuracy. The more flexible loop may experience also further conformations and is essentially disordered. However, this effect cannot be captured by the ECD spectrum which is mostly due to the effect of the stacked guanine nucleobases. This is clearly even more stringent for the simulated spectrum in which the loop is not included in the QM partition. In addition, we have shown that the G4 conformation can, without any major structural rearrangement, form stable complexes with known G4 ligands, which are susceptible to increase the persistence of the quadruplex structure. The important role of G4s in tuning the viral response and the biological cycle, including emerging RNA viruses such as Zika, Dengue, or coronaviruses, calls for the precise determination of putative G4 sequences. Moreover, influencing the equilibrium between unfolded and G4 sequences via the use of small drugs offers an original, yet not fully explored, possibility for the development of novel and

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

potentially wide-action antiviral agents. This study highlights the robustness of our *in silico* protocol to provide a most favorable complement to experimental studies in suggesting specific interaction modes and structures of RNA sequences. Our contribution represents a proof of concept of the capacities offered by mature and multiscale simulation techniques to unravel key biological processes and phenomena.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c03071.

Protocol of the sequence and sequence-structure analysis, additional details of the MD and QM/MM simulations, RNA G4 structural parameters for the different replicas, different force fields, unshifted ECD spectra, 2D-RMSD, and clustering (PDF)

Pdb of RG-1 structure bound and unbound to the ligands (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Giampaolo Barone − *Department of Biological, Chemical and Pharmaceutical Sciences, University of Palermo, 90126 Palermo, Italy;* orcid.org/0000-0001-8773-2359; Email: giampaolo.barone@unipa.it

Antonio Monari − *Université de Lorraine and CNRS, UMR 7019 LPCT, F-54000 Nancy, France; Université de Paris and CNRS, Itodys, F-75006 Paris, France;* orcid.org/0000-0001-9464-1463; Email: Antonio.monari@univ-lorraine.fr

Authors

Tom Miclot − *Department of Biological, Chemical and Pharmaceutical Sciences, University of Palermo, 90126 Palermo, Italy; Université de Lorraine and CNRS, UMR 7019 LPCT, F-54000 Nancy, France*

Cécilia Hognon − *Université de Lorraine and CNRS, UMR 7019 LPCT, F-54000 Nancy, France*

Emmanuelle Bignon − *Université de Lorraine and CNRS, UMR 7019 LPCT, F-54000 Nancy, France;* orcid.org/0000-0001-9475-5049

Alessio Terenzi − *Department of Biological, Chemical and Pharmaceutical Sciences, University of Palermo, 90126 Palermo, Italy;* orcid.org/0000-0001-9751-1373

Marco Marazzi − *Departamento de Química Analítica, Química Física e Ingeniería Química, Universidad de Alcalá, E-28805 Madrid, Spain; Instituto de Investigación Química "Andrés M. del Río" (IQAR), Universidad de Alcalá, E-28871 Madrid, Spain;* orcid.org/0000-0001-7158-7994

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c03071

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work has been conducted via the financial support of the Universities of Palermo and Lorraine and the French CNRS. T.M. thanks the University of Palermo for granting a Ph.D. fellowships. Calculations have been partially performed on the local LPCT computing cluster and on the Regional Explor Computing Center. The authors thank GENCI for providing access to the national computing center under the project "Seek&Destroy". M.M. is grateful to the University of Alcalá for providing funds under the COVID-19 project 2020/00256/001. E.B. and A.M. thank the CNRAS and the French Research Ministry (MESRI) for funding under the GAVO project.

■ REFERENCES

(1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727−733.

(2) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19* (3), 141−154.

(3) Harapan, H.; Itoh, N.; Yufika, A.; Winardi, W.; Keam, S.; Te, H.; Megawati, D.; Hayati, Z.; Wagner, A. L.; Mudatsir, M. Coronavirus Disease 2019 (COVID-19): A Literature Review. *J. Infect. Public Health* **2020**, *13* (5), 667−673.

(4) Forni, G.; Mantovani, A.; Forni, G.; Mantovani, A.; Moretta, L.; Rappuoli, R.; Rezza, G.; Bagnasco, A.; Barsacchi, G.; Bussolati, G.; et al. COVID-19 Vaccines: Where We Stand and Challenges Ahead. *Cell Death Differ.* **2021**, *28* (2), 626−639.

(5) Topol, E. J. Messenger RNA Vaccines against SARS-CoV-2. *Cell* **2021**, *184* (6), 1401.

(6) Beigel, J. H.; Tomashek, K. M.; Dodd, L. E.; Mehta, A. K.; Zingman, B. S.; Kalil, A. C.; Hohmann, E.; Chu, H. Y.; Luetkemeyer, A.; Kline, S.; et al. Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* **2020**, *383* (19), 1813−1826.

(7) Gil, C.; Ginex, T.; Maestro, I.; Nozal, V.; Barrado-Gil, L.; Cuesta-Geijo, M. Á.; Urquiza, J.; Ramírez, D.; Alonso, C.; Campillo, N. E.; et al. COVID-19: Drug Targets and Potential Treatments. *J. Med. Chem.* **2020**, *63* (21), 12359−12386.

(8) Arya, R.; Kumari, S.; Pandey, B.; Mistry, H.; Bihani, S. C.; Das, A.; Prashar, V.; Gupta, G. D.; Panicker, L.; Kumar, M. Structural Insights into SARS-CoV-2 Proteins. *J. Mol. Biol.* **2021**, *433* (2), 166725.

(9) Francés-Monerris, A.; Hognon, C.; Miclot, T.; García-Iriepa, C.; Iriepa, I.; Terenzi, A.; Grandemange, S.; Barone, G.; Marazzi, M.; Monari, A. Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches. *J. Proteome Res.* **2020**, *19* (11), 4291−4315.

(10) Ruggiero, E.; Richter, S. N. Viral G-Quadruplexes: New Frontiers in Virus Pathogenesis and Antiviral Therapy. *Annu. Rep. Med. Chem.* **2020**, *54*, 101−131.

(11) Zaccaria, F.; Paragi, G.; Fonseca Guerra, C. The Role of Alkali Metal Cations in the Stabilization of Guanine Quadruplexes: Why K+ Is the Best. *Phys. Chem. Chem. Phys.* **2016**, *18* (31), 20895−20904.

(12) Fonseca Guerra, C.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M. Telomere Structure and Stability: Covalency in Hydrogen Bonds, Not Resonance Assistance, Causes Cooperativity in Guanine Quartets. *Chem. - Eur. J.* **2011**, *17* (45), 12612−12622.

(13) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS J.* **2010**, *277*, 3459−3469.

(14) Zhao, C.; Song, H.; Scott, P.; Zhao, A.; Tateishi-Karimata, H.; Sugimoto, N.; Ren, J.; Qu, X. Mirror-Image Dependence: Targeting Enantiomeric G-Quadruplex DNA Using Triplex Metallohelices. *Angew. Chem.* **2018**, *130* (48), 15949−15953.

(15) Zhao, C.; Wu, L.; Ren, J.; Xu, Y.; Qu, X. Targeting Human Telomeric Higher-Order DNA: Dimeric G-Quadruplex Units Serve as Preferred Binding Site. *J. Am. Chem. Soc.* **2013**, *135* (50), 18786−18789.

(16) Neidle, S. Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer: G-Quadruplexes as Cancer Drug Targets. *FEBS J.* **2010**, *277* (5), 1118−1125.

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

(17) Abiri, A.; Lavigne, M.; Rezaei, M.; Nikzad, S.; Peyman, Z.; Mergny, J.-L.; Rahimi, H.-R. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol. Rev.* **2021**, *73*, 897−923.
(18) Hognon, C.; Miclot, T.; Garcia-Iriepa, C.; Francés-Monerris, A.; Grandemange, S.; Terenzi, A.; Marazzi, M.; Barone, G.; Monari, A. Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses. *J. Phys. Chem. Lett.* **2020**, *11* (14), 5661−5667.
(19) Tan, J.; Kusov, Y.; Mutschall, D.; Tech, S.; Nagarajan, K.; Hilgenfeld, R.; Schmidt, C. L. The "SARS-Unique Domain" (SUD) of SARS Coronavirus Is an Oligo(G)-Binding Protein. *Biochem. Biophys. Res. Commun.* **2007**, *364* (4), 877−882.
(20) Kusov, Y.; Tan, J.; Alvarez, E.; Enjuanes, L.; Hilgenfeld, R. A G-Quadruplex-Binding Macrodomain within the "SARS-Unique Domain" Is Essential for the Activity of the SARS-Coronavirus Replication−Transcription Complex. *Virology* **2015**, *484*, 313−322.
(21) Ruggiero, E.; Richter, S. N. Survey and Summary G-Quadruplexes and G-Quadruplex Ligands: Targets and Tools in Antiviral Therapy. *Nucleic Acids Res.* **2018**, *46*, 3270−3283.
(22) Yuan, Q.; Wu, Y.; Wang, J.; Lu, D.; Zhao, Z.; Liu, T.; Zhang, X.; Tan, W. Targeted Bioimaging and Photodynamic Therapy Nanoplatform Using an Aptamer-Guided G-Quadruplex DNA Carrier and Near-Infrared Light. *Angew. Chem., Int. Ed.* **2013**, *52* (52), 13965−13969.
(23) Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X. Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chem., Int. Ed.* **2021**, *60* (1), 432−438.
(24) De Nicola, B.; Lech, C. J.; Heddi, B.; Regmi, S.; Frasson, I.; Perrone, R.; Richter, S. N.; Phan, A. T. Structure and Possible Function of a G-Quadruplex in the Long Terminal Repeat of the Proviral HIV-1 Genome. *Nucleic Acids Res.* **2016**, *44* (13), 6442−6451.
(25) Marquevielle, J.; Kumar, M. V. V.; Mergny, J. L.; Salgado, G. F. 1H, 13C, and 15N Chemical Shift Assignments of a G-Quadruplex Forming Sequence within the KRAS Proto-Oncogene Promoter Region. *Biomol. NMR Assignments* **2018**, *12* (1), 123−127.
(26) Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L. A.; Mergny, J. L.; Salgado, G. F. High-Resolution Three-Dimensional NMR Structure of the KRAS Proto-Oncogene Promoter Reveals Key Features of a G-Quadruplex Involved in Transcriptional Regulation. *J. Biol. Chem.* **2017**, *292* (19), 8082−8091.
(27) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954−9960.
(28) Perez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92* (11), 3817−3829.
(29) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886−2902.
(30) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1950.
(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(32) Gattuso, H.; Spinello, A.; Terenzi, A.; Assfeld, X.; Barone, G.; Monari, A. Circular Dichroism of DNA G-Quadruplexes: Combining Modeling and Spectroscopy to Unravel Complex Structures. *J. Phys. Chem. B* **2016**, *120* (12), 3113−3121.
(33) Terenzi, A.; Gattuso, H.; Spinello, A.; Keppler, B. K.; Chipot, C.; Dehez, F.; Barone, G.; Monari, A. Targeting G-Quadruplexes with Organic Dyes: Chelerythrine−DNA Binding Elucidated by Combining Molecular Modeling and Optical Spectroscopy. *Antioxidants* **2019**, *8* (10), 472.
(34) Gattuso, H.; Garcia-Iriepa, C.; Sampedro, D.; Monari, A.; Marazzi, M. Simulating the Electronic Circular Dichroism Spectra of Photoreversible Peptide Conformations. *J. Chem. Theory Comput.* **2017**, *13* (7), 3290−3296.
(35) Neidle, S. Quadruplex Nucleic Acids as Targets for Anticancer Therapeutics. *Nature Reviews Chemistry* **2017**, *1*, 0041.
(36) Tian, T.; Chen, Y. Q.; Wang, S. R.; Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem.* **2018**, *4* (6), 1314−1344.

https://doi.org/10.1021/acs.jpclett.1c03071
J. Phys. Chem. Lett. 2021, 12, 10277−10283

CHAPTER 13

Structural stability of G-quadruplexes

Chapter contents

Understanding the mechanisms affecting the stability of G-quadruplexes is, along with their identification, a subject of great interest. Several studies exist and show different approaches to study the structural stability of G-quadruplexes. They have highlighted the importance of the central ion, which is somehow the keystone of G-quadruplexes since its loss leads to unfolding, and that parallel topology is more easily unfolded than the antiparallel conformation [182]. Other studies have focused on the involvement of adenines in the structure of G-quadruplexes. First, the substitution of an adenine in the quartets does not prevent the formation of a G-quadruplex, but there is a position effect. Indeed, replacing a guanine of the central quartet by an adenine causes the formation of a G-quadruplex structure that is not very stable and bimolecular [183]. It has alsop been shown that the position and the number of adenines present in the loops of G-quadruplexes influence the formation and the stability of the folding [184, 185], and can drive the structural conformation of G-quadruplex [186]. Other studies have focused on the influence of loop size on the structure of G-quadurplexes. They reveal that large-size loop may influences the syn-anti orientation of guanines present in quartets [187]. But also that the convention that the maximum length of 7 bases for loops is overrated. Indeed, Guédin et al. [188] have shown that G-quadruplexes can be formed even in the presence of a loop exceeding 12 nucleotides; and that it is still possible to form a

relatively stable G-quadruplex with a 30 nucleotide loop. Further on this topic, loops do not always appear to be destructured as a single flexible strand. This means that the loops can also adopt a secondary structure: hairpin, and participate in the formation of a stable G-quadruplex [189]. Finally, the stability of G-quadruplexes should also be considered in the framework of DNA lesions, and in particular guanine oxidation and strand breaks. These are the main focus of this chapter.

13.1 Introduction of oxidized guanine into G-quadruplex

Guanines represent the keystone of G-quadruplexes, However, guanine is also the most sensitive nucleotide to reactive oxygen species, its most common oxidation product being 8-oxo-7,8-dihydroguanine (8oxoG) [33, 34, 35]. The introduction of this oxidized form of guanines in G-quadruplexes has been studied. It has been shown that the structural impact is dependent on the position of 8oxoG, but that telomeric G-quadruplexes remain globally stable [190]. The following paper presents a multidisciplinary approach using molecular dynamics calculations in combination with circular dichroism experiments and cellular immunofluorescence assays to assess the resistance of G-qudruplexes to oxidative lesions. The calculations show that the conformation of G-quadruplexes remains globally stable when a single lesion is introduced. Also that the introduction of a double lesion, i.e. of two 8oxoG, gives some unfolded structures even if the majority of the structures remain globally stable. These theoretical results are in perfect agreement with the circular dichroism curves which highlight the resistance of G-quadruplexes to an increasing concentration of hydrogen peroxide. Furthermore, immunofluorescence assays reveal a simultaneous increase between the amount of 8oxoG and G-quadruplex when cells are exposed to hydrogen peroxide.

The introduction of more than one 8oxoG in a G-quadruplex may seem unlike to occur in biological conditions since 8oxoG are even more sensitive to oxidation than guanine. So there should not be several 8oxoG in a G-quadruplex. From a certain point of view, an 8oxoG "protect" the structure from oxidation. However, it is quite possible that

oxidation occurs on the DNA strand before it folds into G-quadruplex. In this case, it is possible to have several 8oxoG in a G-quadruplex, especially if they are distant in the sequence.

Personal contribution My involvement in the paper was the part on molecular dynamics simulations, as well as their analysis. I also carried out all circular dichroism measurements.

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993

Chemistry Europe
European Chemical Societies Publishing

Forever Young: Structural Stability of Telomeric Guanine Quadruplexes in the Presence of Oxidative DNA Lesions**

Tom Miclot,[a, b] Camille Corbier,[c] Alessio Terenzi,[a] Cécilia Hognon,[b] Stéphanie Grandemange,*[c] Giampaolo Barone,*[a] and Antonio Monari*[b]

Abstract: Human telomeric DNA, in G-quadruplex (G4) conformation, is characterized by a remarkable structural stability that confers it the capacity to resist to oxidative stress producing one or even clustered 8-oxoguanine (8oxoG) lesions. We present a combined experimental/computational investigation, by using circular dichroism in aqueous solutions, cellular immunofluorescence assays and molecular dynamics simulations, that identifies the crucial role of the stability of G4s to oxidative lesions, related also to their biological role as inhibitors of telomerase, an enzyme overexpressed in most cancers associated to oxidative stress.

Introduction

DNA G-quadruplexes (G4s) represent a non-canonical nucleic acid arrangement with remarkably different properties compared to the more conventional double-helical B-DNA, first described by Watson and Crick.[1] G4s can form either in DNA and RNA, and they have been correlated to relevant biological effects also related to viral replication.[2–7]

G4s are usually emerging in guanine rich DNA regions, and their most common occurrence is based on a folded single-stranded architecture. The presence of loops connecting the different guanines also allows the formation of G4 between noncontiguous nucleobases, offering a largely increased flexibility.

From a structural point of view, G4s are constituted by a quartet of guanine bases forming planar arrangements (tetrads) and are stabilized by Hoogsteen type hydrogen bonds.[8] These latter non-covalent interactions, have been most notably shown to present a high degree of cooperativity, also justifying the extremely high rigidity of the G4 core.[9–11]

The formation of the tetrad arrangement is accompanied by the accumulation of a quite important negative charge, that could lead to electrostatic repulsion in the center of the quartet, compromising the global stability. For this reason, the central channel is stabilized by the presence of a cation, which leads to a weak electrostatic interaction with the 6-oxygen atoms of the guanines in the tetrad,[12–14] usually monovalent alkaline metal ions such as K^+ and Na^+ are present.[13] Furthermore, low hydration and crowded environments, such as those found in intracellular conditions, are suitable to increase the G4s stability.[15] Indeed, G4s have been recognized as particularly stable and resistant in various conditions. These also include, albeit non exclusively, their capacity to resist to the cleavage by nuclease or the much higher thermal stability compared to other nucleic acid arrangements.

Despite the stability and rigidity of the core structure, G4s can exist in different conformation depending on the relative sugar orientation. Conventionally, this gives rise to the so-called parallel, antiparallel, and hybrid structures,[16,17] obtained by the relative orientations of the backbone and loops connecting the tetrads. The equilibrium between the conformers is also highly sensitive to the environmental conditions and may change drastically depending on the central cation, or on the presence of crowding agents.[18] Recently, we have shown that proper use of a strategy relying on the concomitant use of molecular simulations and spectroscopic techniques, such as electronic circular dichroism (ECD), allows to properly characterize the specific signatures and unequivocally identify G4s topologies.[19,20]

Although the presence of G4s in cellular compartments, either cytoplasmic or nuclear, has been confirmed only rather recently, their biological functions are various and crucial. For instance, they are involved in chromatin remodeling, regulation of replication and gene expression and have been associated with genomic instability, genetic diseases and cancer.[21] Indeed, G4s arrangements are also present in gene promoting regions, allowing a transcriptional regulation of the corresponding gene, as it has been well described for the oncogene *c-myc*. One of the most crucial function of G4s is also to protect the telomeric ends of the chromosomes, that comprise guanine-rich single-

[a] *T. Miclot, Prof. A. Terenzi, Prof. G. Barone*
Department of Biological, Chemical and Pharmaceutical Sciences and Technologies
Università degli Studi di Palermo, Viale delle Scienze, 90128 Palermo (Italy)
E-mail: giampaolo.barone@unipa.it

[b] *T. Miclot, C. Hognon, Prof. A. Monari*
Université de Lorraine and CNRS
LPCT UMR 7019, 54000 Nancy (France)
E-mail: antonio.monari@univ-lorraine.fr

[c] *Dr. C. Corbier, Prof. S. Grandemange*
Université de Lorraine and CNRS
CRAN UMR 7039, 54000 Nancy (France)
E-mail: stephanie.grandemange@univ-lorraine.fr

[**] *A previous version of this manuscript has been deposited on a preprint server (https://www.biorxiv.org/content/10.1101/2020.11.26.399741v1).*

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202100993

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

stranded regions. In this context G4s also act as efficient inhibitors of the telomerase,[22,23] the enzyme controlling telomers length during replication. Since the progressive shortening of telomeres is related to the triggering of cellular senescence and death pathways, its deregulation can be related to carcinogenesis, and in particular to the "immortality" phenotype of cancer cells. As a consequence, G4 stabilizers are nowadays widely considered as potential anticancer drug candidates[24,25] and some of them are presently in clinical trial.[26–29] More generally, maintaining of the G4s structural stability is essential to avoid triggering senescence of the cells and of the organism, especially in the presence of external stress conditions.

As well as other DNA structures, G4s may be subjected to photolesions or oxidative damage that can occur as a consequence of oxidative stress. Indeed, G4s may even be regarded as hot-spots for oxidatively induced lesions since guanine is the most sensitive nucleotide to reactive oxygen species (ROS), and to oxidation in general, since, its reduction potential is the lowest among all the DNA bases.[30] Guanine oxidative products cover a quite large chemical space,[31,32] but 8-oxoguanine (8oxoG) is by far the most common and hence the most characterized lesion.[33–35] 8oxoG may be produced from a direct one electron reaction of hydroxyl radical ($OH^{\bullet}$) on the C8 atom of a guanine, followed by an electron transfer to O_2 and deprotonation.[32] 8oxoG is also produced via DNA photosensitization through the intermediate of singlet oxygen ($^1O_2$).[35,36] Although the structure of 8oxoG is very close to guanine, it has different physicochemical properties, for example, its steric clash is higher. It has also been reported that 8oxoG/C DNA strands have a consistently different hydration environment compared to G/C base pairs in normal DNA.[37] Hence, the introduction of 8oxoGs lesions may have a significant impact on the global structure of DNA.

If the impact of DNA lesions in G4s is less widely analyzed in comparison to that in canonical B-DNA, it has recently emerged that damaged G4s could lead to crucial physicochemical or even biological outcomes that deserve attention. As a matter of fact, Markovitsi's group has recently pointed out that a guanidinium cation, i.e. an intermediate in oxidative lesions pathways, has a much longer life-time in G4s than in canonical DNA.[38] Furthermore, Burrows's group has on the one side identified the presence of 8oxoG in DNA G4s; and on the other side pointed out the specific interaction between the damaged strand and the repair protein machinery.[39] Importantly, it has emerged that the interaction with base excision repair (BER) components may lead to a complex signaling ultimately resulting in the modulation of gene expression and in epigenetic regulation. Other important biological effects have been correlated to the introduction of oxidative lesions inside the G4s sequence, among which the increase of transcription and telomerase activity.[12,40,41] Moreover, the accumulation of oxidative lesions play an important role in both neurodegeneration and carcinogenesis,[42,43] and hence the role of G4s as a promising target for new anticancer drugs should be considered.[4,26,40,44]

In a previous work we have analyzed by molecular modeling and simulation the effects of the presence of apurinic/apyrimidinic (AP) sites on the structure of a telomeric sequence able to form G4 structures (*h-Telo)*.[45] We have pointed out that the G4 stability strongly depends on the position of the AP in the DNA sequence. While usually the damaged G4 structure has been shown to rearrange to maintain the global folded conformation, some cases can be evidenced in which the quadruplex structure is totally unfolded. The unfolding is usually correlated with the disruption of the central leaflet and the concomitant release of both central cations.

Differently from the case of AP, 8oxoGs in G4s usually leads to conformational changes thanks to the recruitment of guanines from the peripheral loop sequence,[46,47] while the specific position of the damage seems to still be critical in dictate the seriousness of the unfolding.[48] The impact of 8oxoG lesions into G4s can be reduced by replacing guanine with xanthine,[49] or by using a pyrene-modified guanine tract.[50]

In the present study we aim to provide a full and systematic analysis on G4s DNA damage,[45] by investigating the structural and biological effects of single and double 8oxoGs lesions. To this aim, we combined molecular modeling and simulation studies with electronic circular dichroism (ECD) spectroscopy of G4s exposed to increasing concentration of hydrogen peroxide (H_2O_2). Furthermore, we performed cellular biology assays to quantify and localize G4s and 8oxoG via immunofluorescence in H_2O_2 treated cell lines. As a note, we may recall that in biological environment telomeric G4s may adopt parallel or hybrid conformation, and a precise disentangling of the conformational space is quite complicated. For this reason, while most of the MD simulations involve parallel strands, we have also checked the results for hybrid configurations. In the same spirit, we also repeated the ECD determination for both conformations obtaining coherent results.

Results and Discussion

Molecular dynamics simulations

We ran 17 simulations, with two replicas, of damaged parallel G4s structures in different orientations (Figure 1). As previously mentioned, control experiments were performed on hybrid structures yielding equivalent results as reported in SI. Globally, the results of MD simulations statistically show a remarkable stability of the G4 structure relative to oxidation. In fact, as summarized in Table 1, the introduction of a single 8oxoG lesion induces negligible changes in the DNA backbone. Only in few cases, ca. 12%, this leads to the expulsion of one central K^+ and eventually to the disruption of one tetrad. However, in most cases, i.e. 75%, the G4 arrangement is preserved (e.g. see Figure 2B). Only a very small number of trajectories (ca. 6%) leads to a completely unfolded DNA conformation (e.g. see Figure 2A).

Not unexpectedly, G4 structural perturbations are more important when two 8oxoG lesions are simultaneously present

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

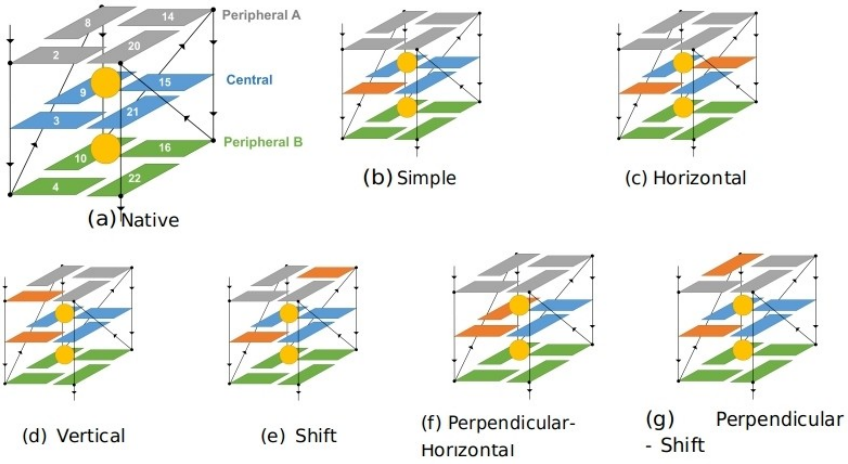


Figure 1. Position of the lesions in the native G4 studied in the present work and the relative orientation of the double lesions. The numbers in panel (a) represent the position of the corresponding guanine in the DNA sequence. Each guanine tetrad has been assigned a color code, i.e. green and grey for the peripheral and blue for the central. 8oxoG in orange.

Table 1. Summary of the main outcome of the MD simulations. (A) Globally conserved: the structure maintains the characteristic of a G4 persistently along the MD; (B) G4-like: only one tetrad is disrupted while the other maintain the G4 arrangement; (C) Disrupted: the G4 structure is totally lost.

	Number of simulation (2 runs)	Number of K+ lost		Topology		
		1	*2*	(A) Globally conserved	(B) G4 like	(C) Disrupted
Parallel						
1 lesion	8	87.50%	12.50%	75.00%	18.75%	6.25%
2 lesions	9	55.56%	44.44%	50.00%	5.56%	44.44%
Hybrid						
1 lesion	2	0.00%	0.00%	100.00%	0.00%	0.00%
2 lesions	2	25.00%	25.00%	75.00%	25.00%	0.00%

in the tetrad (Figure 2 C and D). Note however that double-damaged strands may be representative of situations of high oxidative conditions, in which the structural deformations are extremely pronounced, as observed for DNA cluster lesions.[51,52] Still, with two 8-oxoG lesions our MD simulations show that the G4 conformation is preserved in 50% of total cases considered. This result agrees with the structural stability of G4s in biological conditions and with their recognized protective role against oxidative agents. Remarkably, hybrid G4 conformations are also resistant to the presence of lesions and are conserved in G4 or G4-like arrangements in the presence of either one or two damage site(s) (see Figures S19–S22). The global preservation of the G4 DNA structure is also evidenced by the structural analysis of some snapshots extracted along the MD trajectory. Results are reported in Figure 2 for some representative arrangements and in the SI (Figures S4–S22) for the totality of the trajectories. Note that, since G4 unfolding can be slow, we have also prolonged some of the trajectories up to the 1 μs time limit, without any sign of destabilization.

In addition to the stability of the arrangement, a crucial feature revealed by the MD simulations is the expulsion of one central K^+ cation while preserving the folded quadruplex. In this case, our obtained structures look similar to the stable intermediate described by Zhang et al.[53] The loss of the cation may be a consequence of the structural changes induced by the 8oxoG lesion.

To provide a more quantitative description of the induced deformation we have also analyzed the MD trajectories in terms of crucial structural deformation parameters (Figures S30-S34), following the protocol defined by Tsvetkov et al.[54] in particular, the distance between the G-quartet and the orientation of the guanine was compared to the undamaged G4 strand. As can be seen in Figure S30, the undamaged G4 presents a π-stacking distance of 3.5 Å between the G-quartets, and the twist between the quartets and between the nucleobases inside each plane are also remarkably constant along the MD trajectory. When the G4 structure is lost (e.g. see Figure 2A) we can see from Figure S31 that all the structural signatures are coherently degraded and any pattern is lost. In these conditions the angles between the guanine inside each tetrad also present a remarkable variability. In the case of a structural stable arrangement (situation of Figure 2B) a strikingly different situation can be observed in Figure S32. Indeed, both the stacking distance and the twisting between the quartets remain close to the ideal value showing only slightly increased oscillation. The same stability is also observed for guanine-guanine angles between the tetrads, despite more pronounced oscillations. Globally, the same picture can be sketched for the double-lesioned cases

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

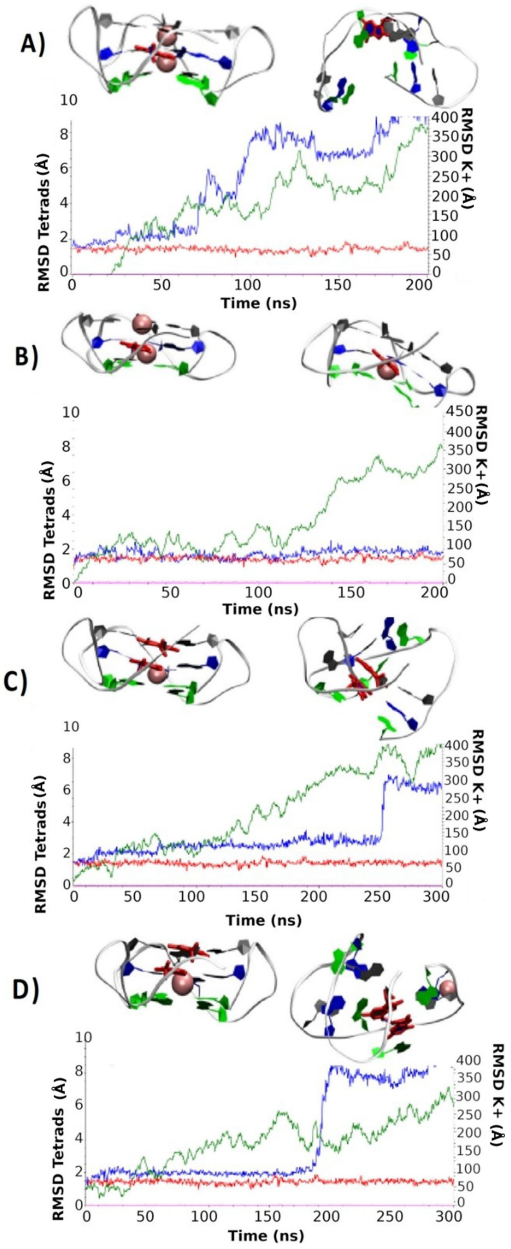


Figure 2. Starting and final conformation obtained for G4s with simple lesion (run 1 A and run 2 B) and vertical double lesions (run 1 C and run 2 D); 8oxoG (in red) is placed at position 3 and 2–3, respectively leading either to conserved or disrupted G4s. The time evolution of the RMSD for the tetrads (red and blue line) and for the central K^+ ions (green and magenta line) are also reported. Note the different scale for the two sets of curves.

(Figure 2C and 2D) reported in Figure S33 and S34. However, in this case the more pronounced structural deformation with a reorganization of the tetrads makes the analysis in terms of global descriptors more complex.

One of the most striking deformations observed concerns the arrangements of guanine O6 in the tetrads. Indeed, the introduction of one 8oxoG changes the configuration of those atoms, which move from a nearly ideal square conformation to a rhomboidal arrangement, as illustrated in Figure 3 and S3, hence perturbing the tetrad shape, without disrupting the global arrangement.

Unsurprisingly, and coherently with what observed for AP sites,[45] the lesions on the central tetrad appears to have a more disruptive effect. However, subtle sequence and position effects should be pointed out. As detailed in Figure 2, we can identify two peculiar situations: when 8oxoG lesions is located on the A peripheral tetrad (position 2 and 14), we may still evidence the structural disturbance of the tetrad while the G4 structure is globally conserved. On the other hand, the lesion in position 16 induces a complete disturbance of the peripheral B tetrad. These observations cannot be explained solely as an effect of steric clash. Indeed, Giorgi et al.[55] point out that an 8oxoG strand is able to assemble in quartet, forming other types of hydrogen bonds than those involving guanines. Hence, 8oxoG is potentially capable of forming non-Hoogsteen hydrogen bonds with a neighboring nucleotide. This is confirmed by our simulations that clearly evidences the formation of persisting non-covalent interactions between 8oxoG and the neighboring guanines involving standard Hoogsteen hydrogen bonds (see Figure S3).

This situation also points to the global conservation of G4-like arrangements. In addition, specific interactions can be pointed out involving either guanine and the cyclopenthenyl moiety of 8oxoG in a fashion already highlighted by Giorgi et al.,[55] and the interactions between H1 and H21 of one nucleotide and the O6 atom of the second partner. Interestingly, this latter interaction, already evidenced experimentally by Bielskutè et al.,[46] is formed not only between undamaged guanine and 8oxoG, but also between two guanines or between two damaged nucleobases (see SI). While the specific outcome of the structural rearrangement strongly depends on the sequence and the position of the 8oxoG lesions, it is clear that the tendency to maintain a G4 or G4-like structure is emerging as the dominant motif when only one lesion is

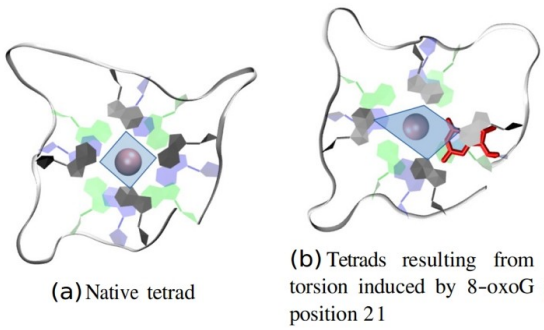


Figure 3. Deformation of the ideal disposition of the tetrad as an effect of the 8oxoG lesion (in red).

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

present, i.e. a situation that can be thought to be compatible with moderate oxidative stress conditions.

The sequence dependence observed in the case of a single 8oxoG lesion is expected to emerge also in the case of a strand featuring two damages. However, as observed in the case of both double strand and G4 DNA in the presence of cluster lesions, the coupling between the two damages may open further deformation paths resulting in different and more extreme structural outcomes. Indeed, as summarized in Table 1, in this case the majority of the MD trajectories leads to globally unfolded structures, that are also accompanied by the concomitant release of both central cations. Interestingly, while 3 trajectories still preserve the G4 arrangement, only in one case the G4-like conformation, with the unfolding of only one tetrad, and the release of only one cation, is observed. Differently from the case of one single lesion, the possible different arrangements of the two cluster lesion, grows combinatorially (Figure 1) including situations in which the damages occupy the same (horizontally arrangement) or different tetrads (vertical arrangement), giving rise to some differences in their coupling and hence in their effects. Unsurprisingly, it turns out that the most disruptive effect is found in the case of horizontally placed lesions, especially when they involve the central tetrad. Indeed, this situation results in the unfold of the first tetrad that is then accompanied with the destabilization of the central K^+, its release and finally the unfolding. On the other hand, in the vertical arrangement one can observe a larger resistance to the lesion. This is also confirmed by the fact that for some of the conformations we obtained two different results for the two replicas, indicating a complex and rather rough free energy landscape that may lead to the coexistence of folded and unfolded structures. Finally, and interestingly, when the lesion occupies both peripheral tetrads, we observe the expulsion of a central cation, and a global reorganization which maintains two of the quartet in a G4-like conformation, made possible by the slight sliding of the remaining K^+ to occupy the region between the two leaflets. This situation is also indicative of a general tendency implying that the 8oxoG containing G4 seems resistant to the loss of one cation, and indeed folded structures in which only one of the cations is present are observed. Furthermore, the unfolding process, in almost all of the cases, is temporally initiated by the loss of a first cation, but necessitates the further expulsion of the second one to be completed. Of particular interest is the case reported in Figure 2D in which we observe two temporally well distinct events: the destabilization of the first tetrad that proceeds rather smoothly and a sharp transition leading to the unfolding of the second one. Snapshots extracted at important time points of the trajectory are reported in Figure 4. Interestingly, the analysis of the trajectory points to a rather complex equilibrium with the bulk K^+ ions that are initially stabilizing the tetrad even after the loss of the two central cations. However, this arrangement leads to a metastable state that rapidly collapses at around 190 ns due, once again to the interaction with the cations that forms a cluster around the G4s and leads to favorable electrostatic interactions with the electron-rich guanines. The cluster of K^+ ions is still present during the first step of the unfolding as shown by the 204 ns conformation.

Thus, our results while globally indicating a much larger structural destabilization produced by double-lesions on the G4 conformation, they also confirm a structural resistance of the G4 arrangement even in the presence of a relative high density of lesions. However, and compared to the presence of AP damages, 8oxoG appears more innocent and even in the case of clustered lesions, i.e. in conditions of high or very high oxidative stress, the structural stability is important, as witnessed by the presence of a non-negligible number of still folded sequences.

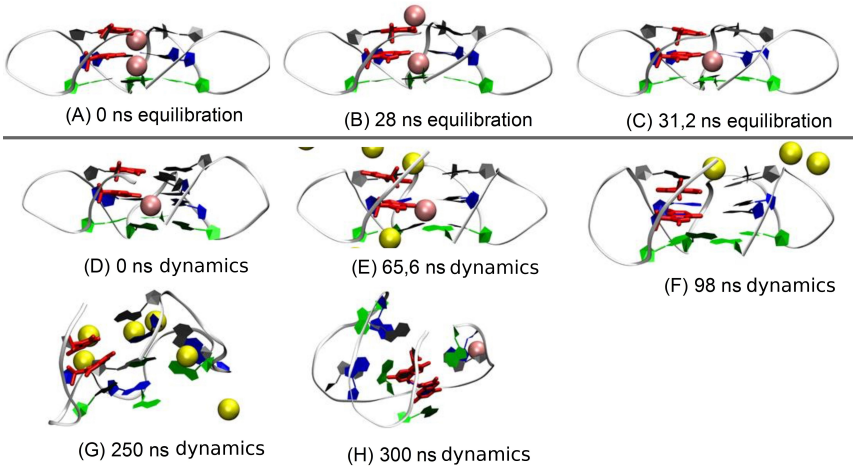


Figure 4. Snapshots representing the dynamic evolution of the double lesioned G4 reported in Figure 2D. 8oxoG in red; crystallographic K^+ in the central canal in pink, bulk K^+ cations in yellow.

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

Circular dichroism

MD simulations are extremely advantageous in the study of the structural reorganizations induced as a consequence of DNA damages, also thanks to the molecular scale resolution that they can offer. Experimentally, a method of choice to unravel the complexity of even subtle structural modifications in biological ordered systems is electronic circular dichroism spectroscopy.[19,56,57] ECD spectra of the *h-Telo* G4 sequence, exposed to increasing concentrations of H_2O_2, in order to induce oxidative damage, were recorded. For the ECD studies, we have considered both a hybrid and a parallel folding, the former obtained in K^+ rich buffered solution and the latter by including a crowding agent, such as poly-ethyleneglycol (PEG-200). The dichroism spectra are reported in Figure 5, while the corresponding UV absorption spectra can be found in Figure S23 and S24.

As can be seen in Figure 5, both arrangements show their typical spectroscopic signatures: a positive band near 265 nm and a negative peak centered around 240 nm for the parallel conformation and two positive peaks at 270 and 290 nm followed by a negative band near 240 nm for the hybrid strands. In both cases, the bands are globally maintained upon addition of H_2O_2. Indeed, while the ECD spectrum for the parallel arrangement is absolutely unchanged upon the addition of oxidant, only a very slight decrease of the intensity of the main positive band can be observed for the hybrid structure. Remarkably, all main spectroscopic features are maintained even in the presence of a 1000-fold excess of H_2O_2. Due to the very strong sensitivity of ECD to secondary structure modifications in biological polymers, the results obtained can be interpreted in terms of a global conservation of the G4 structures and hence of their stability, coherently with the results obtained from the MD simulations. The minor decrease in intensity observed for the hybrid G4s can be attributed to the induction of some structural perturbation following guanine to 8oxoG oxidations that, as shown by the results of MD simulations, should be considered as minor. Of note, the slight perturbation of the ECD and of the UV absorption spectrum reported in Figure S24 can also be seen as an indirect confirmation of the formation of DNA lesions due to the effect of oxidation. A similar experiment was carried out in the presence of copper(II) acetate at 1:1 copper/DNA molar ratio and of H_2O_2 at 100/1 H_2O_2/DNA molar ratio, as indicated by Fleming and Burrows,[58] see Figure S25. Even in these metal-mediated oxidizing condition, the changes in the ECD spectra are negligible, confirming the G4 stability and corroborating the MD results.

In cellulo immunofluorescence

Since the ECD spectra in solution confirm the stability of G4s exposed to oxidative stress, we went a step further in analyzing the behavior of healthy mammary epithelial cell lines, MCF10, exposed to H_2O_2. Via specific immunofluorescence assays, we identified and quantified both the presence of 8oxoG and of G4s. Obviously, and differently from the solution case, in cellular media the amount of G4s after exposition to stress may be related to other factors in addition to the purely structural stability. These may include the influence of repair enzymes, the presence of complex signaling pathways and their cross-talk, and the global cellular response. Furthermore, a dependence upon the cellular cycle may also be observed and pinpointed. Hence, the immunofluorescence assays were also repeated in the presence of antioxidants to assess their effect. As displayed in Figure 6 and SI, we can observe that in absence of oxidative stress a minimal amount of 8oxoG is present, on the other hand the addition of H_2O_2 leads to a very strong and statistically significant increase of its quantity. Unsurprisingly, the concomitant addition of H_2O_2 and antioxidants while still producing a significant amount of DNA lesions almost reduces its increase by half. As for the level of G4s in the presence of antioxidants, H_2O_2 addition clearly leads to an increase that is still statistically significant. However, differently from the 8oxoG content, such increase appears as almost antioxidant independent. The increase of G4 content in condition of oxidative stress has already been documented in different cells lines and should be considered as a defense mechanism of the cell induced by the action of chaperone proteins.[59] As such, it cannot be related uniquely to chemical and structural stability factors. However, the importance of the latter can be seen in the fact that the fluorescent labelling shows a rather large overlap for both 8oxoG and G4s, indicating co-localization. Hence, we may safely

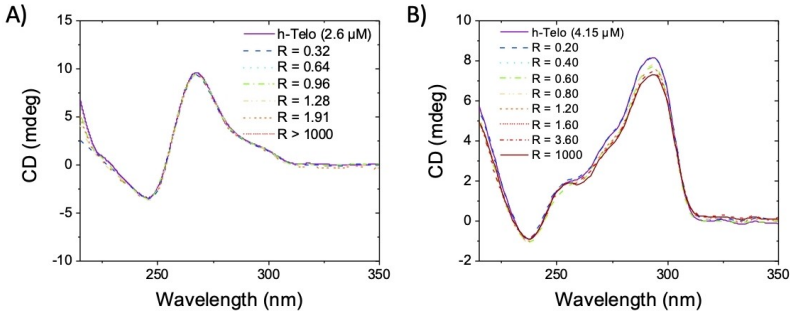


Figure 5. ECD spectra for parallel (A) and hybrid (B) *h-Telo* G4 DNA, recorded in the presence of increasing concentration of H_2O_2 ($R = [H_2O_2]/[h\text{-}Telo]$).

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

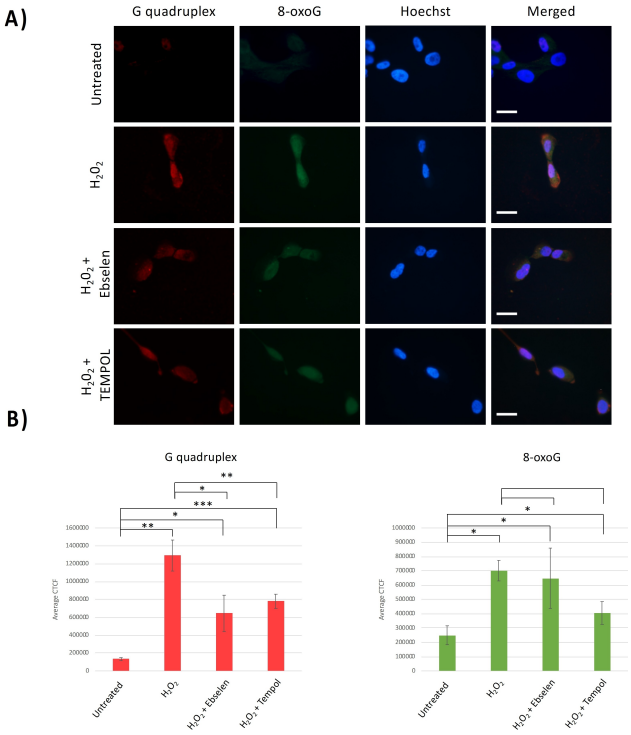


Figure 6. (A) Epifluorescence microscopy of MCF10a cells either untreated, treated with H_2O_2 or H_2O_2 and Ebselen or TEMPOL. Cells were fixed with paraformaldehyde and double stained with G quadruplex (red) and 8-oxoguanine (green) antibodies. Scale bar represents 20 μm. (B) Quantification of fluorescence levels of immunofluorescence on MCF10a cells. Ten cells were counted per experiment. n = 3, T-test: * $p < 0,05$, ** $p < 0,01$.

conclude that while G4s can be seen as hotspots for DNA oxidative lesions, coherently with the high density of guanine, their structural stability helps in maintaining the global arrangement.

Conclusions

The combined use of our multiscale approach has allowed us to clearly sketch the scenario of the G4-DNA behavior in response to oxidative stress producing 8oxoG. All results obtained, going from molecular modeling to cellular biology assays through spectroscopic studies, clearly point out a remarkable structural stability of the G4s to oxidative stress. MD simulations show that almost all the *h-Telo* sequences harboring an isolated damage are stable and do not undergo extended unfolding. Even the inclusion of a secondary lesion, while increasing the amount of unfolded sequences, is still characterized by a higher structurally stability compared to other lesions, such as AP sites. These results nicely support the interpretation of the virtually unchanged ECD spectra observed upon G4 titration with H_2O_2 and the increase of the G4s amount in MCF10 cell lines treated with H_2O_2. Furthermore, the results of the cellular biology assays clearly show global nuclear localization of 8oxoG and G4s, confirming that the chemical structural stability is a prerequisite for the cellular response to oxidative stress, i.e. the increase of the amount of G4s.

The influence of oxidative DNA lesions on the stability of G4s has also been extensively investigated by different authors. For instance, Vorlickova et al.[48] have shown that the position of the lesion is extremely important in determining the stability of the quadruplex, with the central tetrad being the most critical one. Our results, essentially agree with these findings, since MD simulations have clearly shown that lesions at the central tetrad are the most likely to induce structural perturbation or even disruption of the structure. On the same spirit it was observed by Zhou et al.[60] that the oxidation of the central guanine plane was accompanied by more dramatic effects on the G4 CD spectra, even if the authors points to a more important structural deformation as the one observed in the present contribution. As pointed out by Bielskute et al.[46] while some structural deformations are evidenced by substituting guanines with 8oxoG, it also appears that the human telomeric G4 is able to accommodate the damaged nucleobase, especially when in *syn* conformation. Coherently with what observed in the present work, it was established that the damaged G4s still retained structural stability also at physiological pH and temperature.

By using a combination of MD simulations, CD spectroscopy and cellular assays, we have achieved an atomic resolution view

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

of the process by which G4s maintain their structural stability even in the presence of particular lesions. Hence evidencing all the subtle equilibrium between weakened hydrogen bonding network, electrostatic interactions, and π-stacking in determining the resistance of G4 to oxidation.

These results allow us to sketch out some considerations on the biological role of G4-DNA. G4s are in fact known to act as regulators of the gene expression, with the most notable case of the *c-myc* oncogene, or as regulator of telomerase activity. In particular, the inhibition of telomerase by DNA folding in G4 conformation helps to prevent cells to acquire immortality via the progressive shortening of telomeres. In conditions of strong oxidative stress, cells may be exposed to increased amount of DNA damages that may lead to mutations or even carcinogenesis. Hence, it is reasonable that the situation helping to confer immortality, and hence possible aggressivity and tumor-like phenotypes should be avoided. This can be indeed achieved by the stabilization of telomeric G4s. However, G4s are hotspots for oxidative lesions due to their high guanine density, hence a strong structural stability of the DNA folding is required. From a chemical and biophysical point of view this is achieved by the fact that 8oxoG, differently from AP sites, is still able to engage in non-Hoogsteen hydrogen-bonds with neighboring guanines, that while inducing a partial deformation of the tetrads maintain the global arrangement. This in turn is translated also in a much stronger propensity to the maintaining of the central cation, whose stabilizing role is crucial to keep the folded G4. Indeed, the case of AP lesions the global stability was achieved at the expense of one of the quartets, that was sacrificed to lead to a more extensive structural rearrangement. The higher stability is also witnessed by the fact that even the presence of clustered 8oxoG, while obviously inducing a much stronger destabilization of the secondary structure, results in preserved G4 or G4-like conformations.

With our work we have contributed to analyze the effects of oxidative lesions on the structural behavior of G4s, also supporting their biological role. In the future we plan to expand the present study considering the interactions of damaged G4s with protein partners, either BER repair enzymes or transcription factors. The structural effects of other lesions, such as strand breaks in the stability of the G4s arrangements will also be fully considered since the latter can be related to the effects of ionizing radiations.

Experimental Section

Molecular dynamics simulations: In Table 2 we report the description of all the sequences for which MD simulations have been run detailing the position of the 8OxoG in the sequence and in the topological arrangement of the G4. Each MD trajectory has been performed following the same protocol. In all the cases we have considered the *h-Telo* sequence folded in a parallel G4 arrangement (PDB code 1KF1) or a hybrid structure (PDB code 2HY9) as the starting point, to which 8oxoG lesions have been manually added to specific positions. All the damaged DNA sequences have been solvated through a TIP3P box of water.[61] K^+ are added to ensure the electroneutrality of the system; original K^+ cations positioned inside G4 structure are conserved. A buffer of 10 Å of water is added to create the final octahedral box. Standards constants 300 K and 1 atm conditions are used to set up the dynamic simulations in the NPT ensemble. Amber ff99 force field including bsc1 corrections[62] is used to describe DNA, while 8oxoG potential is described by a specific force field designed by Bignon et al. in a previous work.[35] Hydrogen mass repartitioning (HMR)[63] is consistently applied to increase the non-water hydrogen mass, hence allowing the use of a 4 fs time step in combination with the Rattle and Shake algorithms.[64] 1000 step of minimization are

Table 2. Representation of the native and lesioned sequences of hybrid (top) and parallel (bottom) G4 modeled in this study. O: 8OxoG, P_A Peripheral A tetrad, C_e: Central tetrad, P_B peripheral B tetrad.

Lesion type	Orientation	Name	Tetrad position	Sequence 5′–3′
No lesion	/	native	undamaged	AAA GGG TTA GGG TTA GGG TTA GGG AA
8-oxoguanine single		5O	C_e	AAA GOG TTA GGG TTA GGG TTA GGG AA
		18O	P_A	AAA GGG TTA GGG TTA GGO TTA GGG AA
8-oxoguanine double	shift	5–18O s	C_e–P_A	AAA GOG TTA GGG TTA GGO TTA GGG AA
	vertical	4–5O v	P_A–C_e	AAA OOG TTA GGG TTA GGG TTA GGG AA
Lesion type	**Orientation**	**Name**	**Tetrad position**	**Sequence 5′–3′**
no lesion	/	native	undamaged	A GGG TTA GGG TTA GGG TTA GGG
8-oxoguanine single		2O	P_A	A OGG TTA GGG TTA GGG TTA GGG
		3O	C_e	A GOG TTA GGG TTA GGG TTA GGG
		4O	P_B	A GGO TTA GGG TTA GGG TTA GGG
		9O	C_e	A GGG TTA GOG TTA GGG TTA GGG
		14O	P_A	A GGG TTA GGG TTA OGG TTA GGG
		15O	C_e	A GGG TTA GGG TTA GOG TTA GGG
		16O	P_B	A GGG TTA GGG TTA GGO TTA GGG
		21O	C_e	A GGG TTA GGG TTA GGG TTA GOG
8-oxoguanine double	horizontal	2–14O h	P_A	A OGG TTA GGG TTA OGG TTA GGG
		3–15O h	C_e	A GOG TTA GGG TTA GOG TTA GGG
	vertical	2–3O v	P_A–C_e	A OOG TTA GGG TTA GGG TTA GGG
		14–15O v	P_A–C_e	A GGG TTA GGG TTA OOG TTA GGG
	shift	3–14O s	P_A–C_e	A GOG TTA GGG TTA OGG TTA GGG
		4–14O s	P_A–P_B	A GGO TTA GGG TTA OGG TTA GGG
		4–15O s	C_e–P_B	A GGO TTA GGG TTA GOG TTA GGG
	perp.-Hor.	9–3O ph	C_e	A GOG TTA GOG TTA GGG TTA GGG
	perp.-Shi.	9–4O ps	C_e–P_B	A GGO TTA GOG TTA GGG TTA GGG

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

performed on the initial systems to remove bad contacts, followed by equilibration and thermalization for a total of 36 ns. All calculations were performed on two replica and continued until the RMSD of the designed G-quadruplex DNA is stable, i.e. between 200 ns and 300 ns. To avoid artefacts due to the insufficient sampling of the conformational space in case of slow conformational transitions the trajectories for the structures maintaining stable G4 aggregates have been prolonged up to 1 μs, without the appearance of any instability. MD calculations were performed using the NAMD software.[65] All the MD trajectories have been analyzed and visualized using VMD[66] and the corresponding scripts provided by Tsvetkov et al.[54]

Electronic circular dichroism: The *h-Telo* G4 sequence (5'-AGG GTT AGG GTT AGG GTT-3') was purchased from IDT (Integrated DNA Technologies, Belgium) in HPLC purity grade. The oligonucleotide was dissolved in MilliQ water to yield a 100 μM stock solution. This was then diluted using 50 mM Tris-HCl/100 mM KCl buffer (pH 7.4) to the desired concentration. When needed, PEG-200 at 40% w/v was added to the buffer in order to obtain a parallel folding of the G4. The oligonucleotide was annealed heating the solutions up to 90 °C for 5 min and then by slowly cooling down to room temperature overnight; *h-Telo* concentration was checked measuring the absorbance at 260 nm and using 184000 L/(mol·cm) as extinction coefficient. Hydrogen peroxide concentration was determined by a redox titration with $KMnO_4$. A stock solution of 0.982 M of H_2O_2 was kept in the fridge and used fresh for the *h-Telo* oxidation experiment. The ECD titrations were carried out by adding increasing amounts of properly diluted H_2O_2 to a solution of *h-Telo* at fixed concentration. Each measurement was performed after 5 min after mixing the two solutions in the cuvette. Furthermore, we have repeated the last measurement after 30 min. To ensure the occurrence of guanine oxidation in the presence of H_2O_2, a similar experiment was carried out in the presence of copper(II) acetate at 1:1 copper/DNA molar ratio and of H_2O_2 at 100/1 H_2O_2/DNA molar ration, as indicated by Fleming and Burrows[58] (see Fig. S25).

Immunofluorescence assays: Immunofluorescence assays were performed on MCF10a cells, a non-tumorigenic human breast epithelial cell line. MCF10a cells were cultured at 37 °C, 5% CO_2 in DMEM/F12, supplemented with 5% horse serum, 2 mM L-glutamine, 100 U/mL penicillin, 100 U/mL streptomycin, 10 μg/mL bovine insulin, 0,5 μg/mL hydrocortisone, 100 ng/mL cholera toxin and 20 ng/mL hEGF. MCF10a cells were cultured on a glass slide in complete medium for 24 h before treatment. Cells were treated with either 200 μM H_2O_2 (H1009, Sigma-Aldrich) for 1 h with or without antioxidant, 50 μM Ebselen (E3520, Sigma Aldrich) or 3 mM TEMPOL (**176141**, **Sigma Aldrich)**. Cells were fixed in 4% paraformaldehyde over 20 minutes at room temperature, then blocked and permeabilized with PBS-containing 2% BSA/0.2% Triton X-100. Cells were incubated with anti-DNA/RNA G-quadruplex [BG4] primary antibody (Ab00174-1.1, Absolute Antibody) diluted at 1:200 in PBS-containing 2% BSA for 30 minutes at 37 °C. After three washes in PBS, the cells were incubated with Alexa Fluor 594 -conjugated goat anti mouse secondary antibody (A-11032, Invitrogen), diluted at 1:500 in PBS-containing 2% BSA for 20 minutes at 37 °C. The cells were washed three times in the washing buffer 1X of OxyDNA Assay Kit (500095, Calbiochem) and incubated overnight at 37 °C in FITC-conjugated anti 8-oxoguanine (500095, Calbiochem) diluted at 1:100 in washing buffer 1X. After three washes with washing buffer, nuclei were stained with Hoechst diluted at 1:10,000 in PBS. The cells were then mounted in antifading medium (FluorSafe; Merck) and observed with an epifluorescence microscope Eclipse 80i with ×100 oil immersion objective (Nikon). Images were collected with a digital camera (Nikon, DS–Ri1) with the same exposure time for all the conditions. Cells fluorescence levels have been assessed following measuring cell fluorescence using ImageJ entry in The Open Lab Book, contributed by Luke Hammond, QBI, The University of Queensland, Australia Hammond 48. Cells were selected on ImageJ using the drawing tool, and their area, integrated density and mean grey value were measured. Regions around the cell were selected and measured to determine the background. Corrected total cell fluorescence (CTCF) was determined as the integrated density minus the product between the area of the selected cell and the mean fluorescence of the background reading. Graphs were made using mean CTCF value for each condition, error bars correspond to SEM.

Acknowledgements

Support from the Ministero dell'Università e Ricerca Scientifica e Tecnologica and Università di Palermo, Italy, and Université de Lorraine and CNRS, France, are gratefully acknowledged. Tom Miclot thanks University of Palermo for funding a joint Ph.D. program. CINECA – SCAI national computing center is acknowledged for graciously providing access to computational resource. Part of the calculations have been performed on the local LPCT computational resources and on the regional ExpLor mesocenter in the framework of the project "Dancing Under the Light".

Conflict of Interest

The authors declare no conflict of interest.

Keywords: electronic circular dichroism · Guanine quadruplexes · immunofluorescence · molecular dynamics · oxidative DNA lesions

[1] J. D. Watson, F. H. C. C. Crick, *Nature* **1953**, *171*, 737–738.
[2] A. M. Fleming, Y. Ding, A. Alenko, C. J. Burrows, *ACS Infect. Dis.* **2016**, *2*, 674–681.
[3] C. Hognon, T. Miclot, C. G. Iriepa, A. Francés-Monerris, S. Grandemange, A. Terenzi, M. Marazzi, G. Barone, A. Monari, *J. Phys. Chem. Lett.* **2020**, *11*, 5661–5667.
[4] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.
[5] J. F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Mégnin-Chanet, C. Hélène, J. L. Mergny, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2672–2677.
[6] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
[7] H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
[8] J. Moon, J. H. Han, D. Y. Kim, M. Jung, K. Kim, *Biochem. Biophys. Reports* **2015**, *2*, 29–35.
[9] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128.
[10] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *Chem. Eur. J.* **1999**, *5*, 3581–3594.
[11] C. Fonseca Guerra, H. Zijlstra, G. Paragi, F. M. Bickelhaupt, *Chem. Eur. J.* **2011**, *17*, 12612–12622.
[12] J. Spiegel, S. Adhikari, *Trends Chem.* **2020**, *2*, 123–136.
[13] F. Zaccaria, G. Paragi, C. Fonseca Guerra, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20895–20904.
[14] C. Nieuwland, F. Zaccaria, C. Fonseca Guerra, *Phys. Chem. Chem. Phys.* **2020**, *22*, 21108–21118.

Chemistry–A European Journal | Full Paper doi.org/10.1002/chem.202100993 | Chemistry Europe European Chemical Societies Publishing

[15] M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane, J. O. Trent, *J. Am. Chem. Soc.* **2010**, *132*, 17105–17107.
[16] Y. Ma, K. Iida, K. Nagasawa, *Biochem. Biophys. Res. Commun.* **2020**, *531*, 3–17.
[17] A. M. Varizhuk, A. D. Protopopova, V. B. Tsvetkov, N. A. Barinov, V. V. Podgorsky, M. V. Tankevich, M. A. Vlasenok, V. V. Severov, I. P. Smirnov, E. V. Dubrovin, D. V. Klinov, E. Pozmogova, *Nucleic Acids Res.* **2018**, *46*, 8978–8992.
[18] J. Li, J. J. Correia, L. Wang, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, 4649–4659.
[19] H. Gattuso, A. Spinello, A. Terenzi, X. Assfeld, G. Barone, A. Monari, *J. Phys. Chem. B* **2016**, *120*, 3113–3121.
[20] A. Biancardi, A. Burgalassi, A. Terenzi, A. Spinello, G. Barone, T. Biver, B. Mennucci, *Chem. Eur. J.* **2014**, *20*, 7439–7447.
[21] P. Prorok, M. Artufel, A. Aze, P. Coulombe, I. Peiffer, L. Lacroix, A. Guédin, J.-L. Mergny, J. Damaschke, A. Schepers, C. Cayrou, M.-P. Teulade-Fichou, B. Ballester, M. Méchali, *Nat. Commun.* **2019**, *10*, 3274.
[22] Y. Xue, Z. Y. Kan, Q. Wang, Y. Yao, J. Liu, Y. H. Hao, Z. Tan, *J. Am. Chem. Soc.* **2007**, *129*, 11185–11191.
[23] J. W. Shay, W. E. Wright, *Nat. Rev. Genet.* **2019**, *20*, 299–309.
[24] C. Ducani, G. Bernardinelli, B. Högberg, B. K. Keppler, A. Terenzi, *J. Am. Chem. Soc.* **2019**, *141*, 10205–10213.
[25] R. Bonsignore, F. Russo, A. Terenzi, A. Spinello, A. Lauria, G. Gennaro, A. M. Almerico, B. K. Keppler, G. Barone, *J. Inorg. Biochem.* **2018**, *178*, 106–114.
[26] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
[27] T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong, L. Q. Gu, *ChemMedChem* **2008**, 690–713.
[28] A. Terenzi, R. Bonsignore, A. Spinello, C. Gentile, A. Martorana, C. Ducani, B. Högberg, A. M. Almerico, A. Lauria, G. Barone, *RSC Adv.* **2014**, *4*, 33245–33256.
[29] H. Xu, M. Di Antonio, S. McKinney, V. Mathew, B. Ho, N. J. O'Neil, N. Dos Santos, J. Silvester, V. Wei, J. Garcia, et al., *Nat. Commun.* **2017**, *8*, 1–18.
[30] S. Steenken, S. V. Jovanovic, *J. Am. Chem. Soc.* **1997**, *119*, 617–618.
[31] W. L. Neeley, J. M. Essigmann, *Chem. Res. Toxicol.* **2006**, *19*, 491–505.
[32] G. Pratviel, B. Meunier, *Chem. Eur. J.* **2006**, *12*, 6018–6030.
[33] J. L. Ravanat, R. J. Turesky, E. Gremaud, L. J. Trudel, R. H. Stadler, *Chem. Res. Toxicol.* **1995**, *8*, 1039–1045.
[34] E. Dumont, R. Grüber, E. Bignon, C. Morell, J. Aranda, J.-L. L. Ravanat, I. Tuñón, *Chem. Eur. J.* **2016**, *22*, 12358–12362.
[35] E. Dumont, R. Grüber, E. Bignon, C. Morell, Y. Moreau, A. Monari, J. L. Ravanat, *Nucleic Acids Res.* **2016**, *44*, 56–62.
[36] J. Cadet, T. Douki, D. Gasparutto, J. L. Ravanat, *Mutat. Res.* **2003**, *531*, 5–23.
[37] S. K. Singh, M. W. Szulik, M. Ganguly, I. Khutsishvili, M. P. Stone, L. A. Marky, B. Gold, *Nucleic Acids Res.* **2011**, *39*, 6789–6801.
[38] A. Banyasz, L. Martínez-Fernández, C. Balty, M. Perron, T. Douki, R. Improta, D. Markovitsi, *J. Am. Chem. Soc.* **2017**, *139*, 10561–10568.
[39] A. M. Fleming, J. Zhu, S. A. Howpay Manage, C. J. Burrows, *J. Am. Chem. Soc.* **2019**, *141*, 11036–11049.
[40] C. K. Kwok, C. J. Merrick, *Trends Biotechnol.* **2017**, *35*, 997–1013.
[41] S. C. J. Redstone, A. M. Fleming, C. J. Burrows, *Chem. Res. Toxicol.* **2019**, *32*, 437–446.
[42] Y. Nakabeppu, *Int. J. Mol. Sci.* **2014**, *15*, 12543–12557.
[43] Z. Sheng, S. Oka, D. Tsuchimoto, N. Abolhassani, H. Nomaru, K. Sakumi, H. Yamada, Y. Nakabeppu, *J. Clin. Invest.* **2012**, *122*, 4344–4361.
[44] V. Đapić, V. Abdomerović, R. Marrington, J. Peberdy, A. Rodger, J. O. Trent, P. J. Bates, *Nucleic Acids Res.* **2003**, *31*, 2097–2107.
[45] C. Hognon, A. Gebus, G. Barone, A. Monari, *Antioxidants* **2019**, *8*, 337.
[46] S. Bielskutė, J. Plavec, P. Podbevšek, *J. Am. Chem. Soc.* **2019**, *141*, 2594 2603.
[47] C. A. Omaga, A. M. Fleming, C. J. Burrows, *Biochemistry* **2018**, *57*, 2958–2970.
[48] M. Vorlícková, M. Tomasko, A. J. Sagi, K. Bednarova, J. Sagi, *FEBS J.* **2012**, *279*, 29–39.
[49] V. V. Cheong, B. Heddi, C. J. Lech, A. T. Phan, *Nucleic Acids Res.* **2015**, *43*, 10506–10514.
[50] S. Takahashi, K. T. Kim, P. Podbevšek, J. Plavec, B. H. Kim, N. Sugimoto, *J. Am. Chem. Soc.* **2018**, *140*, 5774–5783.
[51] E. Bignon, H. Gattuso, C. Morell, F. Dehez, A. G. Georgakilas, A. Monari, E. Dumont, *Nucleic Acids Res.* **2016**, *44*, 8588–8599.
[52] H. Gattuso, E. Durand, E. Bignon, C. Morell, A. G. Georgakilas, E. Dumont, C. Chipot, F. Dehez, A. Monari, *J. Phys. Chem. Lett.* **2016**, *7*, 3760–3765.
[53] M. L. Zhang, Y. P. Xu, A. Kumar, Y. Zhang, W. Q. Wu, *Biochemistry* **2019**, *58*, 3955–3959.
[54] V. Tsvetkov, G. Pozmogova, A. Varizhuk, *J. Biomol. Struct. Dyn.* **2016**, *34*, 705–715.
[55] T. Giorgi, S. Lena, P. Mariani, M. A. Cremonini, S. Masiero, S. Pieraccini, J. P. Rabe, P. Samori, G. P. Spada, G. Gottarelli, *J. Am. Chem. Soc.* **2003**, *125*, 14741–14749.
[56] A. Terenzi, H. Gattuso, A. Spinello, B. K. Keppler, C. Chipot, F. Dehez, G. Barone, A. Monari, *Antioxidants* **2019**, *8*, 472.
[57] H. Gattuso, C. Garcia-Iriepa, D. Sampedro, A. Monari, M. Marazzi, *J. Chem. Theory Comput.* 3290–3296.
[58] A. M. Fleming, C. J. Burrows, *Chem. Res. Toxicol.* **2013**, *26*, 593–607.
[59] A. K. Byrd, B. L. Zybailov, L. Maddukuri, J. Gao, J. C. Marecki, M. Jaiswal, M. R. Bell, W. C. Griffin, M. R. Reed, S. Chib, S. G. Mackintosh, A. M. MacNicol, G. Baldini, R. L. Eoff, K. D. Raney, *J. Biol. Chem.* **2016**, *291*, 18041–18057.
[60] J. Zhou, A. M. Fleming, A. M. Averill, C. J. Burrows, S. S. Wallace, *Nucleic Acids Res.* **2015**, *43*, 4039–4054.
[61] P. Mark, L. Nilsson, *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
[62] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, T. E. Cheatham, *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.
[63] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.
[64] S. Miyamoto, Kollman, P. A. Settle, *J. Comput. Chem.* **1992**, *13*, 952–962.
[65] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.
[66] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33-38, 27–28.

Manuscript received: March 18, 2021
Accepted manuscript online: April 19, 2021
Version of record online: May 13, 2021

13.2 G-quadruplexes resistances to strand break damage

Strand breaks consist into the cleavage of the phosphodiester bond $-\mathrm{O}-\mathrm{PO}_2-\mathrm{O}-$ of DNA backbone. This lesion can be produced by ionizing radiation [27], oxidative stress [191, 192], or enzymes [193]. In fact, there are several types of strand breaks and they are treated by different repair pathways [194]. On this subject, the bacterium *Deinococcus radiodurans* shows a remarkable resistance to DNA damage induced by ionizing radiation. First of all, it possesses a protein named DdrC which is able to colocalize at the level of multiple damage sites located on a single DNA strand [195]. This is not the only resistance strategy employed by the bacterium, as it possesses the ability to repair non-canonical forms of strand breaks ending in $5' - \mathrm{OH}/3' - \mathrm{PO}_4^-$ ends [196]. Further into the main topic of this thesis, the resistance of *D. radiodurans* also involves the presence of G-quadruplex in its genome [52]. Concerning the impact of ionizing radiation on G-quadruplexes we can point to the work of Kumari et al. [197], in which it is shown that damage is localized in the G-quadrplexes. When a strand break damage takes place in a loop, it causes it to open. But their work highlights the resistance of G-quadruplexes core to ionizing radiation and tends towards the idea that they provide radiation protection. The effect of ionizing radiation and stand breaks on the stability of G-quadruplexes is poorly documented, when one looks for a description at the atomic level. In this context, the following paper proposes a structural study on the effect of introducing one, or more, strand breaks. It is known that this type of lesions can produce two types of canonical $5' - \mathrm{PO}_4^-/3' - \mathrm{OH}$, and non-canonical $5' - \mathrm{OH}/3' - \mathrm{PO}_4^-$ ends [28]. Thus using molecular dynamics simulations, the study proposed here shows that telomeric G-quadruplex DNAs are resistant to canonical or non-canonical strand break, even when present at the quartet level

molecules

MDPI

Article

Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage

Tom Miclot [1,2], Cécilia Hognon [2], Emmanuelle Bignon [2], Alessio Terenzi [1], Stéphanie Grandemange [3], Giampaolo Barone [1,*] and Antonio Monari [4,*]

[1] Department of Biological, Chemical and Pharmaceutical Sciences, University of Palermo, viale delle Scienze, Ed. 17, 90128 Palermo, Italy; tom.miclot@unipa.it (T.M.); alessio.terenzi@unipa.it (A.T.)
[2] Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France; cecilia.hognon@univ-lorraine.fr (C.H.); emmanuelle.bignon@univ-lorraine.fr (E.B.)
[3] Université de Lorraine and CNRS, CRAN UMR 7039, F-54000 Nancy, France; stephanie.grandemange@univ-lorraine.fr
[4] Université Paris Cité and CNRS, ITODYS, F-75006 Paris, France
* Correspondence: giampaolo.barone@unipa.it (G.B.); antonio.monari@u-paris.fr (A.M.)

Abstract: DNA integrity is an important factor that assures genome stability and, more generally, the viability of cells and organisms. In the presence of DNA damage, the normal cell cycle is perturbed when cells activate their repair processes. Although efficient, the repair system is not always able to ensure complete restoration of gene integrity. In these cases, mutations not only may occur, but the accumulation of lesions can either lead to carcinogenesis or reach a threshold that induces apoptosis and programmed cell death. Among the different types of DNA lesions, strand breaks produced by ionizing radiation are the most toxic due to the inherent difficultly of repair, which may lead to genomic instability. In this article we show, by using classical molecular simulation techniques, that compared to canonical double-helical B-DNA, guanine-quadruplex (G4) arrangements show remarkable structural stability, even in the presence of two strand breaks. Since G4-DNA is recognized for its regulatory roles in cell senescence and gene expression, including oncogenes, this stability may be related to an evolutionary cellular response aimed at minimizing the effects of ionizing radiation.

Keywords: guanine quadruplexes; DNA strand breaks; molecular modeling and simulation

Citation: Miclot, T.; Hognon, C.; Bignon, E.; Terenzi, A.; Grandemange, S.; Barone, G.; Monari, A. Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage. *Molecules* **2022**, *27*, 3256. https://doi.org/10.3390/molecules27103256

Academic Editor: Iztok Turel

Received: 27 April 2022
Accepted: 18 May 2022
Published: 19 May 2022

1. Introduction

Guanine-quadruplex (G4) DNA or RNA structures can be produced by both intra-strand (i.e., produced by the folding of a single-stranded DNA fragment) or inter-strand G-G pairing and may adopt various topologies, depending on the orientation of the glycosidic bond, giving rise to parallel, antiparallel, and hybrid arrangements [1]. The rigid tetrad cores are connected by nucleotide loops, whose length and flexibility may exhibit rather large variations. The structural and dynamic properties of G4s have important and versatile biological implications. For example, DNA or RNA G4s may regulate viral infection cycles, becoming, as a consequence, potential therapeutic targets for antiviral drug candidates [2,3]. For this reason, they are also of major interest in the context of the current pandemic caused by the infectious pathogen SARS-CoV-2, whose genome has been shown to contain G4-compatible regions [4–8]. Interesting and recent reviews on the antiviral possibilities offered by G4 can be found in literature [9–11]. G4s are also involved in some neurological diseases, such as alpha-thalassemia or X-linked intellectual disability syndrome, in which they are positively or negatively involved in a cascade of gene expression regulation [12–14]. G4s are also involved in DNA replication pathways [15,16] and in gene expression, since they have been localized in oncogene and viral DNA promoter regions [17–20]. In addition, G4 are abundant in the terminal sequences of chromosomes,

Molecules **2022**, *27*, 3256

the telomeres, playing an important role in regulating the cellular life cycle by controlling replication-induced shortening of the telomeres, and hence, cellular programmed death via inhibition of telomerase [21–24]. As a matter of fact, the disruption of this mechanisms is linked to the immortality phenotype of cancer cells, which makes G4s ideal targets for cancer chemotherapeutic agents. Remarkably, G4s are also involved in conferring *Deinococcus radiodurans* its extraordinary resistance to ionizing radiation [25,26].

Because of their versatile and rather ubiquitous biological roles, several studies have focused on the effect of different DNA damage on the stability of G4s. In particular, oxidative damage has been scrutinized due to the fact that G4s are inherently composed of guanine-rich sequences, and the latter is the most easily oxidized nucleotide, and 8-oxo-guanine (8-OxoG) is its most common oxidation product [27–30]. Although G4s are clearly considered hotspots for oxidative DNA damage, they have shown strong structural resistance to this class of lesion [31], depending on the amount of oxidative lesions and their position in the DNA backbone [31–34].

Besides guanine modification or deletion, oxidative stress [35,36] and ionizing radiation [37] are also able to induce DNA strand-break damage. Strand breaks may occur at two different points of the same DNA strand, leading to a very strong genome instability, usually resulting in cell death. The resulting highly toxic damage is essentially due to the difficulty in repairing the dispersed DNA fragments. Interestingly, radiation-resistant bacteria possess specific DNA binding proteins that colocalize at the lesion foci favoring, in this way, their repair [38]. Ionizing radiation can result in two kinds of strand-breaks: canonical 5′-PO_4^-/3′-OH (CA), and non-canonical 5′-OH/3′-PO_4^- (NC) [39,40], as shown in Figure 1. Both kinds of DNA terminations are known to occur in biological systems, produced by the activity of deoxyribonucleases I and II [41], respectively. However, the two types of damages are not equivalent: CA strand breaks are also commonly produced by normal cellular processes, notably during replication and repair [42], and thus are more easily recognized by DNA ligase, which may catalyze the formation of a phosphodiester bond to repair the DNA damage [43,44]. Conversely, NC lesions are hardly repaired within the cell, and are usually produced by deoxyribonuclease II during programmed cell death pathways [45,46].

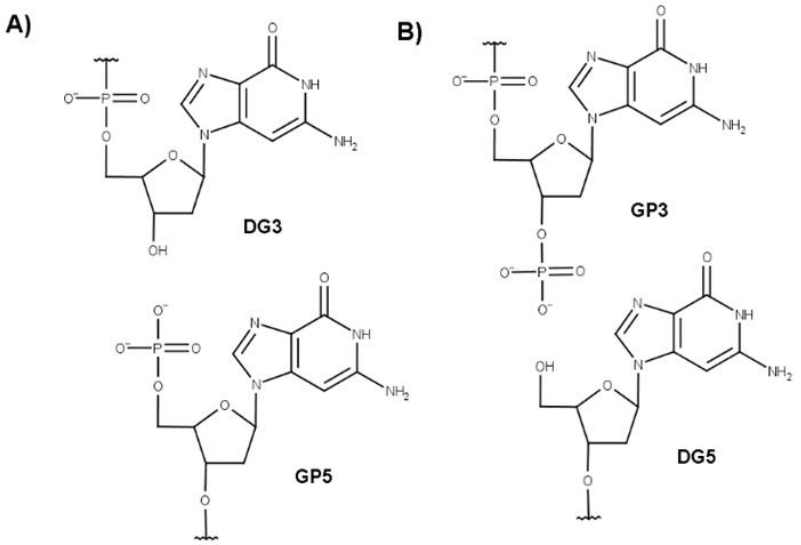


Figure 1. (**A**) Canonical and (**B**) non-canonical strand break damages occurring in the phosphodiester -O-P(O2)-O- bond of a nucleic acid.

However, some pathways allowing the reparation of NC strand breaks exist. As an example, mammalian polynucleotide kinase adds a phosphate group at the 5′ position of non-canonical terminations while replacing the -PO_4^- moiety at the 3′ position with a

hydroxyl group, hence permitting the further action of DNA ligase [47,48]. In addition, some organisms have developed proper mechanisms for repairing non-canonical strand breaks. One example is the repair pathway involving RNA ligase RtcB in *Escherichia coli* or HD-Pnk in *Deinococcus radiodurans* [49,50].

The presence of G4s may induce strand break formation but also may oppose resistance to strand breaks. For example, if G4s cannot be unfolded by helicases, DNA replication is stopped and DNA breaks may occur at their location [51,52]. Conversely, they are also known to play an important role in radio-resistance and the response to DNA damage [25,53]. Recently, Kumari et al. [53] experimentally demonstrated the resistance of G4s to ionizing radiation and their presence in coding DNA sequence (CDS), although this resistance could also be ascribed to a shielding effect of the DNA backbone. Despite their interesting features, the impact of the occurrence of strand break lesions in G4s have been poorly documented from a structural and atomistic point of view.

Here, we investigate the structural effects of the presence of strand breaks in G4-forming DNA sequences containing both CA and NC lesions using state-of-the-art molecular modelling and simulation techniques. In particular, molecular dynamics (MD) simulations have been performed to check how the number and position of strand breaks affect the structure of a parallel G4 structure with a human telomeric sequence (h-telo). Our results show that G4 structures are extremely resistant to strand-breaks, an occurrence which may be correlated to a possible protective role exerted in conditions of high ionizing or oxidative stress.

2. Results

A parallel h-telo G4 DNA was used as our model due to its presence in cells where it protects telomeres by acting as a telomerase inhibitor [54]. Although the h-telo sequence is characterized by high polymorphism, hybrid conformations are also possible [55], the parallel arrangement represents a suitable model for assessing stability in the presence of strand break lesions. Furthermore, we have previously shown that hybrid sequences show even higher stability to oxidative lesions than parallel arrangements [31]. In addition to the undamaged quadruplex used as a reference, 27 different structures harboring strand breaks have been taken into account. For both CA and NC forms, six single-breaks and six double-breaks have been introduced into the tetrads. Breaks, in CA forms, have also been introduced into the loops at three different positions, as shown in Table 1 and Figure 2. This choice allowed us to investigate the role of the position of the strand breaks on the structural stability of G4s. Note that the chosen strand break pattern follows a similar scheme used by us to study the impact of 8-oxoG on the structural stability of G4s [56]. We would also like to emphasize that in the present contribution, the word double-break refers to the presence of two cuts in a quadruplex sequence. Hence, it should not be confused with the usual double-strand nomenclature for B-DNA that refers to nearby breaks in two complementary strands.

Table 1. Schematic representation of the positions of the considered strand breaks. as indicated by the symbol ★. CA and NC strand breaks occur on the two O-P(O2)-O sides of the same phosphate group.

Strand Break Type	Strand Break Position	Sequence
Native		A GGG TTA GGG TTA GGG TTA GGG
Single	2–3	A G★GG TTA GGG TTA GGG TTA GGG
	3–4	A GG★G TTA GGG TTA GGG TTA GGG
	9–10	A GGG TTA GG★G TTA GGG TTA GGG
	14–15	A GGG TTA GGG TTA G★GG TTA GGG
	15–16	A GGG TTA GGG TTA GG★G TTA GGG
	21–22	A GGG TTA GGG TTA GGG TTA GG★G

Table 1. *Cont.*

Strand Break Type	Strand Break Position	Sequence
Double	2–3–4	A G★G★G TTA GGG TTA GGG TTA GGG
	2–3/14–15	A G★GG TTA GGG TTA G★GG TTA GGG
	3–4/9–10	A GG★G TTA GG★G TTA GGG TTA GGG
	3–4/14–15	A GG★G TTA GGG TTA G★GG TTA GGG
	3–4/15–16	A GG★G TTA GGG TTA GG★G TTA GGG
	14–15–16	A GGG TTA GGG TTA G★G★G TTA GGG

The symbol ★ indicates the position of the CA and NC strand breaks occurring on the two O-P(O2)-O sides of the same phosphate group.

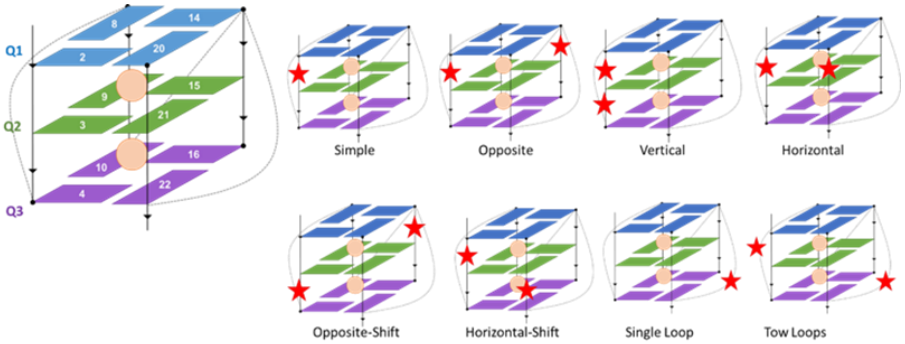


Figure 2. Position of tetrads in the studied h-telo G4 and the relative orientation of the strand break damage (displayed as red stars) in peripheral loops or the tetrad-connecting backbone. The orange dots represent cations.

The presence of strand breaks inevitably induces some structural variations impacting the G4, which may be more or less marked, highly localized, or conversely, more global. Monitoring the evolution of the main structural parameters allowed us to account for the impact of the type of lesion and its position on the specific G4. Namely, we focused on the distance for the centers of mass of the tetrads, their twist angle, and the angles formed between guanines belonging to the same quartet. In addition, the RMSD of the guanines forming the tetrads highlights conservation of the quartet arrangement and global preservation of the G4 conformation.

As shown in the Supplementary Materials (Figures S1–S56) and in Figure 3, RMSD clearly highlights conservation of the quartet arrangement of the G4, and hence the stability of this conformation. This is particularly visible in the 2D-RMSD maps plotted for each structure (Figures 3 and S1–S56). Indeed, the structural deviation of the quartets remained relatively small throughout the simulation, rarely exceeding 3 Å. However, when considering the whole nucleic acid structure, we observed much larger structural variations, even for the native G4. This result is due to the peripheral loops, whose flexibility was clearly enhanced by the presence of strand breaks and constitutes further evidence for the coexistence of a rigid core with flexible loops in G4 arrangements. Interestingly, as shown in Figure 3 for specific strand break positions, this effect is similar and is produced analogously for both CA and NC structures. Remarkably, the stability of the G4 arrangement was also preserved in presence of multiple strand breaks, in contrast to what is observed for other lesions, such as abasic sites [56]. This is of great importance since ionizing radiation deposits its energy in a limited spatial area, usually leading to DNA cluster lesions [57,58].

The behavior of G4 is in striking contrast to canonical double-helical B-DNA structures. Indeed, in the latter case, a double-strand break induces high instability of the genome, which is mainly due to the dispersion of broken DNA fragments. Conversely, G4s appear to be much more resilient and strand breaks do not alter its arrangement. This feature can be correlated with the biological role of G4s and the protective role they can play in the presence of high oxidative stress. The global stability of G4s despite strand breaks also

Molecules **2022**, *27*, 3256

resonates with their resilience in the presence of oxidative damage, which we have recently determined [31]. However, it has to be pointed out that the guanine core may be much more sensitive to damage. Indeed, the introduction of an abasic site [56] may lead to G4 disruption or complex structural reorganization necessary to maintain its folding.

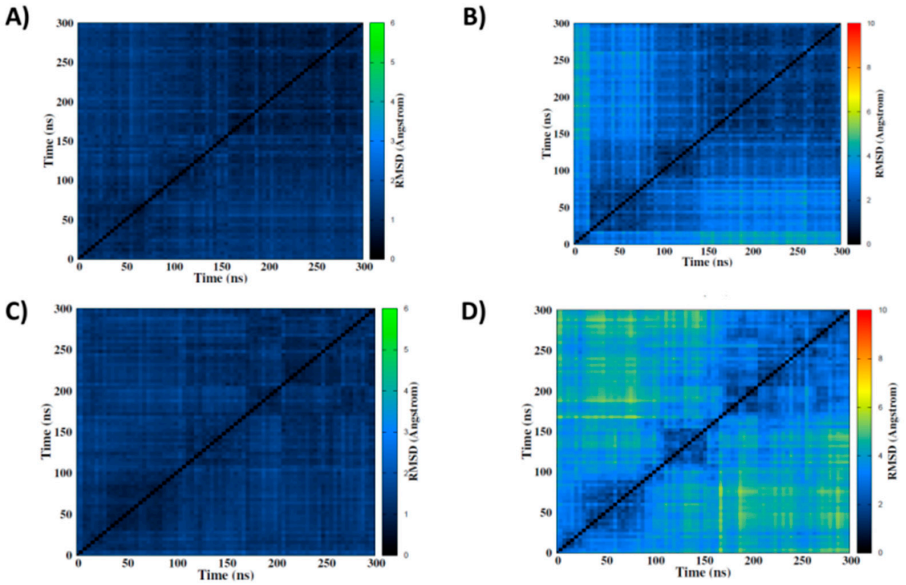


Figure 3. 2D-RMSD maps of simulations of the native structure CA (top) and the 14-15-16 NC (bottom) including the tetrads only (**A**,**C**) or the whole DNA (**B**,**D**).

Our results agree with those reported by Kumari et al. [53], who experimentally demonstrated the resistance of G4s to strand breaks. Their in vitro experiments highlighted the formation of stable intra- and intermolecular G4s after exposure to ionizing radiation. In addition, their cell irradiation experiments suggested that G4-forming regions also exhibited high ionizing radiation resistance. Furthermore, their experimental results pointed to the fact that strand breaks occur mainly in G4-connecting loops. In fact, the introduction of a strand-break lesion into the loops caused them to open [53].

Such a situation undoubtedly leads to a modification of the global arrangement of G4 DNA. The phenomenon is particularly well highlighted by our 2D-RMSD maps (see Figures S1–S56). However, as the guanine core is not affected, the structural properties of the tetrads remained unchanged. Interestingly, from our simulations we may infer that strand breaks located in the connecting loops show an even higher structural resistance compared to those directly connecting the guanines that form the internal core.

To provide a deeper analysis of the effects of strand break formation on G4 structures, we also considered more local deformations. As a matter of fact, Hoogsteen base-pairing of guanines and the interaction with alkali metal ions are crucial for the formation of G4s [59,60], whereas the involvement of the backbone should be considered minor for dictating their formation.

The time series for the angles formed between the guanines in the tetrads still show a global stability (see Figure S57 and Table S1). More precisely, the values were all globally centered around the values measured in the native undamaged structure, i.e., ca. 90° for adjacent and ca. 180° for opposite guanines, with fluctuations being indicative of only slightly minor changes in their arrangement. If we focus on the respective distances between the center-of-mass of the tetrads, we observe only slight fluctuations compared to the native structure, in agreement with the global stability revealed by the RMSD analysis.

Instead, larger fluctuations were be observed for the twist angles (see Table S1 and Figure 4). Indeed, the presence of strand breaks induced a slight enlargement of their

distribution, which may be related to increased flexibility. This is particularly evident for the twist angle between the two terminal tetrads, whereas variations in the twists involving the central tetrad were less important. In addition to enlargement of the distribution, we also noticed some deviations in the maximum value that may lead to a deviation between 5 and 10° from the undamaged structure, pointing to a slightly, albeit non-negligible, structural reorganization. This deviation was maximal when the lesion was in the backbone directly connecting two tetrads, and less pronounced when the loops were involved.

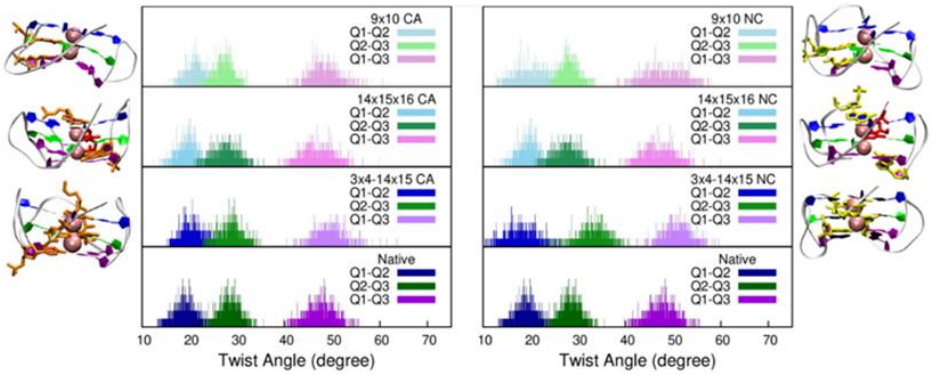


Figure 4. Distribution of the twist angle for representative DNA damaged structures (upper panels) compared to those of the native structure (bottom panels).

Finally, and despite the global stability observed and discussed, we should point out that a significant exception was observed for one trajectory. This involved one replica for CA damage at the 14–15 position (Figure S10). In this case, we observed rearrangement of the first peripheral quartet that was accompanied by the leakage of a K^+ cation, leading to the expulsion of the guanine G8 from a loop region and complete destabilization of the peripheral tetrad. In the past, we have observed that loss of the cation is an important phenomenon in the destabilization of the G4 structure [31,56]. Although this case highlights that metal cations are crucial role for the stability of G4 arrangements, it an remains isolated instance, since it was only observed once in all of our simulations and can therefore be considered a rare event.

3. Discussion and Conclusions

Strand breaks are important DNA lesions typically produced by exposure to ionizing radiation. Double-breaks in a DNA strand are difficult to repair and are associated with very high cytotoxicity. In this contribution, we investigated the impact of strand breaks on the stability and persistence of intra-strand G4 architecture, considering the important regulatory role played by G4-DNA structures at the cellular level. Although double-helical B-DNA strand breaks are usually correlated with strong structural destabilization and consequent genome dispersion, our results consistently showed that G4s experienced only negligible structural deformation in 26 out of 27 cases and maintained their global folding and shape. The introduction of a strand break is typically accompanied by a slight increase in flexibility of the connecting loops, and a slight change of the twist angle, especially when the break is directly located between guanines belonging to different tetrads. Only one strand break, i.e., between the 14th and 15th nucleotides, led to unfolding of the G4 in the second replica. This was due to the deformation of one tetrad and subsequent expulsion of the stabilizing K^+ cation, which can be considered a rare event. Our results, which agree with those reported by Kumari et al. [53], confirm the high stability of G4s and their inherent resistance to strand-break damage. From a molecular point of view, this can be attributed to the combined effect of intra-tetrad Hoogsteen H-bonds (conferring rigidity to the tetrad core) and inter-tetrad π-stacking interactions (contributing to the maintenance of global G4 folding). From a cellular point of view, this occurrence can be related to the

protective and regulatory role played by G4s in regulating gene expression and cellular senescence. As already noted for oxidative lesions, preserving G4 arrangement during conditions of high stress has the effect of limiting the expression of oncogenes, inhibiting telomerase, and avoiding the emergence of an immortal phenotype, i.e., phenomena that can lead to carcinogenesis.

It is important to point out that our MD simulations consistently used a pre-folded G4 on top of which strand breaks were created. Hence, our results are unequivocal concerning the stability of the quadruplexes. However, we could not infer whether the presence of strand breaks could inhibit the folding itself, and hence, the presence of actual G4 in cells. The propensity for G4 folding of damaged DNA, which might also strongly depend on the length of the fragments, and hence, radiation intensity, will require enhanced sampling procedures that will be addressed in forthcoming contributions.

4. Materials and Methods

4.1. Force Field for Non-Standard Nucleotides

Prior to modelling the formation of strand breaks, specific force field parameters for the two sides of the cleaved backbone needed to be generated. To this aim, we chose the AMBER ff99bsc1 force field [61,62] and specific modifications to the guanine and adenine force field were performed to obtain CA and NC ends (see Supplementary Materials). The geometry of each of the new residues was optimized at the B3LYP/6-311+G(d,p) level of theory, with Gaussian 09 [63]. Restrained electrostatic potential (RESP) charges were obtained at HF/6-31G * and converted into amber format with the antechamber utilities.

4.2. Molecular Dynamics Simulations

All simulations were generated using the AMBER16 suite of programs [64]. The initial G4 structure of h-telo was obtained from the pdb database (PDB:1KF1) [65], and the strand breaks were manually created at specific sequence positions (see Table 1 and ESI). Then, the initial systems were solvated in an octahedral TIP3P water box [66] with a 12 Å buffer, and electroneutrality was provided by the addition of K^+ ions. Note that the central K^+ ions present in the crystal structure have been kept. Hydrogen mass repartitioning [67] was applied to allow the integration of the Newton equations of motion using a 4 fs time step in combination with the RATTLE and SHAKE algorithms [68]. All MD simulations were performed with NAMD software [69,70] until a simulation time of 300 ns was reached in the NPT ensemble maintained by a Langevin thermostat and barostat [71]. Each of simulation was preceded by 1000 minimization steps and 36 ns of equilibration. All simulations were performed on two independent replicas to increase global sampling. VMD [72] was used to visualize and analyze the MD trajectories. G-quadruplex structural parameters were calculated using the script developed by Tsvetkov et al. [73].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/molecules27103256/s1, full methodological details and structural analysis of the different G4 strands.

Author Contributions: Simulations, T.M. and C.H.; conceptualization, A.M., G.B., E.B., A.T. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Trajectories of the MD simulations are available upon request.

Acknowledgments: Support from the Universities of Lorraine, Palermo, and Paris Cité, as well as from the French CNRS are gratefully acknowledged. T.M. thanks Universita degli Studi di Palermo for funding his doctoral program. EB thanks the French "Ministère de l'Education Supérieure, de la Recherche Scientifique et de l'Innovation" (MESRI) for funding her postdoctoral fellowship through the special GAVO program. AM thanks ANR (Agence Nationale de la Recherche) and CGI (Commissariat à l'Investissement d'Avenir) for their financial support of this work through Labex SEAM (Science and Engineering for Advanced Materials and devices) ANR 11 LABX 086, ANR 11 IDEX 05 02. Support from the IdEx "Université Paris 2019" ANR-18-IDEX-0001, from the Platform PSM3B, and from European Union 2014-2020 PON "Ricerca e Innovazione" grant from the Italian Ministry of Education, University and Research, "PROGEMA" (ARS01_00432) are gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Burge, S.; Parkinson, G.N.; Hazel, P.; Todd, A.K.; Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415. [CrossRef] [PubMed]
2. Majee, P.; Kumar Mishra, S.; Pandya, N.; Shankar, U.; Pasadi, S.; Muniyappa, K.; Nayak, D.; Kumar, A. Identification and characterization of two conserved G-quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. *Sci. Rep.* **2020**, *10*, 1477. [CrossRef] [PubMed]
3. Metifiot, M.; Amrane, S.; Litvak, S.; Andreola, M.L. G-quadruplexes in viruses: Function and potential therapeutic applications. *Nucleic Acids Res.* **2014**, *42*, 12352–12366. [CrossRef]
4. Hognon, C.C.; Miclot, T.; Iriepa, C.G.; Francés-Monerris, A.; Grandemange, S.; Terenzi, A.; Marazzi, M.; Barone, G.; Monari, A.; Garcia-Iriepa, C.; et al. Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses. *J. Phys. Chem. Lett.* **2020**, *11*, 5661–5667. [CrossRef] [PubMed]
5. Panera, N.; Tozzi, A.; Alisi, A. The G-quadruplex/helicase world as a potential antiviral approach against COVID-19. *Drugs* **2020**, *80*, 941–946. [CrossRef] [PubMed]
6. Tan, J.; Vonrhein, C.; Smart, O.S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J.R.; Schmidt, C.L.; Hilgenfeld, R. The SARS-Unique Domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* **2009**, *5*, e1000428. [CrossRef] [PubMed]
7. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X. Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chem. Int. Ed.* **2021**, *60*, 432–438. [CrossRef]
8. Miclot, T.; Hognon, C.; Bignon, E.; Terenzi, A.; Marazzi, M.; Barone, G.; Monari, A. Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses. *J. Phys. Chem. Lett.* **2021**, *12*, 10277–10283. [CrossRef]
9. Abiri, A.; Lavigne, M.; Rezaei, M.M.; Nikzad, S.S.; Zare, P.; Mergny, J.-L.L.; Rahimi, H.-R.R.; Peyman, Z.; Mergny, J.-L.L.; Rahimi, H.-R.R. Unlocking G-quadruplexes as antiviral targets. *Pharmacol. Rev.* **2021**, *73*, 897–923. [CrossRef]
10. Ruggiero, E.; Richter, S.N. Survey and summary G-quadruplexes and G-quadruplex ligands: Targets and tools in antiviral therapy. *Nucleic Acids Res.* **2018**, *46*, 3270–3283. [CrossRef]
11. Ruggiero, E.; Zanin, I.; Terreri, M.; Richter, S.N. G-quadruplex targeting in the fight against viruses: An update. *Int. J. Mol. Sci.* **2021**, *22*, 10984. [CrossRef] [PubMed]
12. Asamitsu, S.; Takeuchi, M.; Ikenoshita, S.; Imai, Y.; Kashiwagi, H.; Shioda, N. Perspectives for applying g-quadruplex structures in neurobiology and neuropharmacology. *Int. J. Mol. Sci.* **2019**, *20*, 2884. [CrossRef] [PubMed]
13. Shioda, N.; Yabuki, Y.; Yamaguchi, K.; Onozato, M.; Li, Y.; Kurosawa, K.; Tanabe, H.; Okamoto, N.; Era, T.; Sugiyama, H.; et al. Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome. *Nat. Med.* **2018**, *24*, 802–813. [CrossRef] [PubMed]
14. Wang, E.; Thombre, R.; Shah, Y.; Latanich, R.; Wang, J. G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Res.* **2021**, *49*, 4816–4830. [CrossRef]
15. Paeschke, K.; Capra, J.A.; Zakian, V.A. DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase. *Cell* **2011**, *145*, 678–691. [CrossRef]
16. Prorok, P.; Artufel, M.; Aze, A.; Coulombe, P.; Peiffer, I.; Lacroix, L.; Guédin, A.; Mergny, J.-L.; Damaschke, J.; Schepers, A.; et al. Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.* **2019**, *10*, 3274. [CrossRef]
17. Yuan, L.; Tian, T.; Chen, Y.; Yan, S.; Xing, X.; Zhang, Z.; Zhai, Q.; Xu, L.; Wang, S.; Weng, X.; et al. Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy. *Sci. Rep.* **2013**, *3*, 01811. [CrossRef]
18. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 2105. [CrossRef]
19. Lago, S.; Nadai, M.; Cernilogar, F.M.; Kazerani, M.; Domíniguez Moreno, H.; Schotta, G.; Richter, S.N. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **2021**, *12*, 3885. [CrossRef]

20. Robinson, J.; Raguseo, F.; Nuccio, S.P.; Liano, D.; Di Antonio, M. DNA G-quadruplex structures: More than simple roadblocks to transcription? *Nucleic Acids Res.* **2021**, *49*, 8419–8431. [CrossRef]
21. Xue, Y.; Kan, Z.Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y.H.; Tan, Z. Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. *J. Am. Chem. Soc.* **2007**, *129*, 11185–11191. [CrossRef] [PubMed]
22. Okamoto, K.; Seimiya, H. Revisiting Telomere Shortening in Cancer. *Cells* **2019**, *8*, 107. [CrossRef] [PubMed]
23. Neidle, S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer: G-quadruplexes as cancer drug targets. *FEBS J.* **2010**, *277*, 1118–1125. [CrossRef] [PubMed]
24. Wu, R.A.; Upton, H.E.; Vogan, J.M.; Collins, K. Telomerase Mechanism of Telomere Synthesis. *Annu. Rev. Biochem.* **2017**, *86*, 439–460. [CrossRef]
25. Mishra, S.; Chaudhary, R.; Singh, S.; Kota, S.; Misra, H.S. Guanine quadruplex DNA regulates gamma radiation response of genome functions in the radioresistant bacterium deinococcus radiodurans. *J. Bacteriol.* **2019**, *201*, e00154. [CrossRef]
26. Saranathan, N.; Vivekanandan, P. G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* **2019**, *27*, 148–163. [CrossRef]
27. Pratviel, G.; Meunier, B. Guanine oxidation: One- and two-electron reactions. *Chem. A Eur. J.* **2006**, *12*, 6018–6030. [CrossRef]
28. Steenken, S.; Jovanovic, S.V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* **1997**, *119*, 617–618. [CrossRef]
29. Cadet, J.; Douki, T.; Ravanat, J.L. Oxidatively generated damage to cellular DNA by UVB and UVA radiation. *Photochem. Photobiol.* **2015**, *91*, 140–155. [CrossRef]
30. Fleming, A.M.; Burrows, C.J. Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. *J. Am. Chem. Soc.* **2020**, *142*, 1115–1136. [CrossRef]
31. Miclot, T.; Corbier, C.; Terenzi, A.; Hognon, C.; Grandemange, S.; Barone, G.; Monari, A. Forever Young: Structural Stability of Telomeric Guanine Quadruplexes in the Presence of Oxidative DNA Lesions **. *Chem. A Eur. J.* **2021**, *27*, 8865–8874. [CrossRef] [PubMed]
32. Bielskute, S.; Plavec, J.; Podbevšek, P. Impact of Oxidative Lesions on the Human Telomeric G-Quadruplex. *J. Am. Chem. Soc.* **2019**, *141*, 2594–2603. [CrossRef] [PubMed]
33. Szalai, V.A.; Singer, M.J.; Thorp, H.H. Site-specific probing of oxidative reactivity and telomerase function using 7,8-dihydro-8-oxoguanine in telomeric DNA. *J. Am. Chem. Soc.* **2002**, *124*, 1625–1631. [CrossRef] [PubMed]
34. Vorlícková, M.; Tomasko, M.; Sagi, A.J.; Bednarova, K.; Sagi, J. 8-Oxoguanine in a quadruplex of the human telomere DNA sequence. *FEBS J.* **2012**, *279*, 29–39. [CrossRef] [PubMed]
35. Burrows, C.J.; Muller, J.G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **1998**, *98*, 1109–1151. [CrossRef]
36. Driessens, N.; Versteyhe, S.; Ghaddhab, C.; Burniat, A.; De Deken, X.; Van Sande, J.; Dumont, J.E.; Miot, F.; Corvilain, B. Hydrogen peroxide induces DNA single- and double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ. *Endocr. Relat. Cancer* **2009**, *16*, 845–856. [CrossRef]
37. Henner, W.D.; Rodriguez, L.O.; Hecht, S.M.; Haseltine, W.A. gamma Ray induced deoxyribonucleic acid strand breaks. 3′ Glycolate termini. *J. Biol. Chem.* **1983**, *258*, 711–713. [CrossRef]
38. Banneville, A.-S.; Tour, C.B.d.l.; Hognon, C.; Colletier, J.-P.; Teulon, J.-M.; Roy, A.L.; Pellequer, J.-L.; Monari, A.; Dehez, F.; Confalonieri, F.; et al. Structural and functional characterization of DdrC, a novel DNA damage-induced nucleoid associated protein involved in DNA compaction. *bioRxiv* **2021**. [CrossRef]
39. Hennertg, W.; Grunbergsl, S.; Haseltinell, W. Sites and Structure of Gamma Radiation-induced DNA Strand Breaks. *J. Biol. Chem.* **1982**, *257*, 11750–11754. [CrossRef]
40. Von Sonntag, C.; Hagen, U.; Schön-Bopp, A.; Schulte-Frohlinde, D. Radiation-Induced Strand Breaks in DNA: Chemical and Enzymatic Analysis of End Groups and Mechanistic Aspects. *Adv. Radiat. Biol.* **1981**, *9*, 109–142.
41. Lauková, L.; Konečná, B.; Janovičová, L.; Vlková, B.; Celec, P. Deoxyribonucleases and their applications in biomedicine. *Biomolecules* **2020**, *10*, 1036. [CrossRef] [PubMed]
42. Mehta, A.; Haber, J.E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016428. [CrossRef] [PubMed]
43. Tomkinson, A.E.; Della-Maria, J.A. DNA Ligases: Mechanism and Functions. In *Encyclopedia of Biological Chemistry*, 2nd ed.; Academic Press: London, UK, 2013; pp. 28–32, ISBN 9780123786319.
44. Tomkinson, A.E.; Vijayakumar, S.; Pascal, J.M.; Ellenberger, T. DNA ligases: Structure, reaction mechanism, and function. *Chem. Rev.* **2006**, *106*, 687–699. [CrossRef] [PubMed]
45. Didenko, V.V. 5′OH DNA breaks in apoptosis and their labeling by topoisomerase-based approach. *Methods Mol. Biol.* **2011**, *682*, 77–87.
46. Saito, Y.; Hikita, H.; Nozaki, Y.; Kai, Y.; Makino, Y.; Nakabori, T.; Tanaka, S.; Yamada, R.; Shigekawa, M.; Kodama, T.; et al. DNase II activated by the mitochondrial apoptotic pathway regulates RIP1-dependent non-apoptotic hepatocyte death via the TLR9/IFN-β signaling pathway. *Cell Death Differ.* **2019**, *26*, 470–486. [CrossRef]
47. Chappell, C.; Hanakahi, L.A.; Karimi-Busheri, F.; Weinfeld, M.; West, S.C. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J.* **2002**, *21*, 2827–2832. [CrossRef]

48. Karimi-Busheri, F.; Lee, J.; Tomkinson, A.E.; Weinfeld, M. Repair of DNA strand gaps and nicks containing 3′-phosphate and 5′-hydroxyl termini by purified mammalian enzymes. *Nucleic Acids Res.* **1998**, *26*, 4395–4400. [CrossRef]
49. Das, U.; Chakravarty, A.K.; Remus, B.S.; Shuman, S. Rewriting the rules for end joining via enzymatic splicing of DNA 3′-PO4 and 5′-OH ends. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20437–20442. [CrossRef]
50. Schmier, B.J.; Shuman, S. Deinococcus radiodurans HD-Pnk, a nucleic acid end-healing enzyme, abets resistance to killing by ionizing radiation and mitomycin C. *J. Bacteriol.* **2018**, *200*, e00151. [CrossRef]
51. Obi, I.; Rentoft, M.; Singh, V.; Jamroskovic, J.; Chand, K.; Chorell, E.; Westerlund, F.; Sabouri, N. Stabilization of G-quadruplex DNA structures in Schizosaccharomyces pombe causes single-strand DNA lesions and impedes DNA replication. *Nucleic Acids Res.* **2020**, *48*, 10998–11015. [CrossRef]
52. van Kregten, M.; Tijsterman, M. The repair of G-quadruplex-induced DNA damage. *Exp. Cell Res.* **2014**, *329*, 178–183. [CrossRef] [PubMed]
53. Kumari, N.; Vartak, S.V.; Dahal, S.; Kumari, S.; Desai, S.S.; Gopalakrishnan, V.; Choudhary, B.; Raghavan, S.C. G-quadruplex Structures Contribute to Differential Radiosensitivity of the Human Genome. *iScience* **2019**, *21*, 288–307. [CrossRef] [PubMed]
54. Shay, J.W.; Wright, W.E. Telomeres and telomerase: Three decades of progress. *Nat. Rev. Genet.* **2019**, *20*, 299–309. [CrossRef] [PubMed]
55. Rocca, R.; Palazzesi, F.; Amato, J.; Costa, G.; Ortuso, F.; Pagano, B.; Randazzo, A.; Novellino, E.; Alcaro, S.; Moraca, F.; et al. Folding intermediate states of the parallel human telomeric G-quadruplex DNA explored using Well-Tempered Metadynamics. *Sci. Rep.* **2020**, *10*, 3176. [CrossRef] [PubMed]
56. Hognon, C.; Gebus, A.; Barone, G.; Monari, A. Human DNA telomeres in presence of oxidative lesions: The crucial role of electrostatic interactions on the stability of guanine quadruplexes. *Antioxidants* **2019**, *8*, 337. [CrossRef] [PubMed]
57. Bignon, E.; Gattuso, H.; Morell, C.; Dehez, F.; Georgakilas, A.G.; Monari, A.; Dumont, E. Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion. *Nucleic Acids Res.* **2016**, *44*, 8588–8599. [CrossRef]
58. Gattuso, H.; Durand, E.; Bignon, E.; Morell, C.; Georgakilas, A.G.; Dumont, E.; Chipot, C.; Dehez, F.; Monari, A. Repair Rate of Clustered Abasic DNA Lesions by Human Endonuclease: Molecular Bases of Sequence Specificity. *J. Phys. Chem. Lett.* **2016**, *7*, 3760–3765. [CrossRef]
59. Moon, J.; Han, J.H.; Kim, D.Y.; Jung, M.; Joon, K.; Kim, S.; Kim, S.K. Effects of deficient of the Hoogsteen base-pairs on the {G}-quadruplex stabilization and binding mode of a cationic porphyrin. *Biochem. Biophys. Rep.* **2015**, *2*, 29–35. [CrossRef]
60. Largy, E.; Mergny, J.L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met. Ions Life Sci.* **2016**, *16*, 203–258.
61. Galindo-Murillo, R.; Robertson, J.C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T.E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. [CrossRef]
62. Dans, P.D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M. How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **2017**, *45*, 4217–4230. [CrossRef] [PubMed]
63. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 09. Gaussian 09 Revis. D.01*; Gaussian, Inc.: Wallingford, CT, USA, 2009. Available online: https://gaussian.com/glossary/g09/ (accessed on 25 April 2022).
64. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [CrossRef] [PubMed]
65. Parkinson, G.; Lee, M.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Exp. Mol. Pathol.* **2002**, *417*, 876–880. [CrossRef] [PubMed]
66. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
67. Hopkins, C.W.; Le Grand, S.; Walker, R.C.; Roitberg, A.E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874. [CrossRef]
68. Miyamoto, S.; Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962. [CrossRef]
69. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [CrossRef]
70. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [CrossRef]
71. Feller, S.E.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621. [CrossRef]
72. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]
73. Tsvetkov, V.; Pozmogova, G.; Varizhuk, A. The systematic approach to describing conformational rearrangements in G-quadruplexes. *J. Biomol. Struct. Dyn.* **2016**, *34*, 705–715. [CrossRef] [PubMed]

5

G-quadruplex interactions with proteins

CHAPTER 14

G-quadruplexes can promote the dimerization of proteins.

The article presented here is part of our mobilization at the beginning of the 2020 global pandemic. The work focuses on the mechanism of infection of the SARS-CoV-2 virus. More precisely, the SARS Unique Domain (SUD) is a dimeric protein composed of two symmetrical monomers. It interacts with the RNA of the host cell, also in the form of G-quadruplex [198, 199]. The role of the SUD protein is to inhibit the apoptosis signaling pathway and hence promote cell survival, which leads to increased viral replication and evasion from the immune system [198]. Thus, SUD is an important topic of interest, since understanding the mechanisms of its interaction with G-quadruplexes may provide leads for the design of therapeutic molecules. The research presented below gives a structural explanation at the molecular level of these mechanisms, and in particular the role of G-quadrupolexes in promoting the active dimeric form. The analysis of molecular dynamics simulations reveals two possible modes of interaction. One is an interaction between G-quadruplex and a single monomer of the protein, the other is an interaction between G-quadruplex at the monomer/monomer interface. Furthermore, free energy calculations show that the dimeric interaction favors the active form of the protein.

Personal contribution I created the models of the two starting complexes between G-quadruplex and the SUD protein and performed the molecular dynamics simulations, carrying out a series of structural analysis on the trajectories.

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL Letter

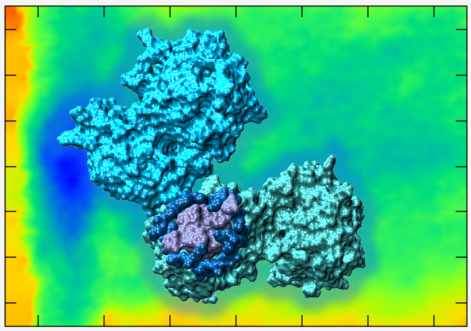
Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses

Cécilia Hognon,[▽] Tom Miclot,[▽] Cristina García-Iriepa, Antonio Francés-Monerris, Stephanie Grandemange, Alessio Terenzi, Marco Marazzi,* Giampaolo Barone,* and Antonio Monari*

Cite This: J. Phys. Chem. Lett. 2020, 11, 5661−5667 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Coronaviruses may produce severe acute respiratory syndrome (SARS). As a matter of fact, a new SARS-type virus, SARS-CoV-2, is responsible for the global pandemic in 2020 with unprecedented sanitary and economic consequences for most countries. In the present contribution we study, by all-atom equilibrium and enhanced sampling molecular dynamics simulations, the interaction between the SARS Unique Domain and RNA guanine quadruplexes, a process involved in eluding the defensive response of the host thus favoring viral infection of human cells. Our results evidence two stable binding modes involving an interaction site spanning either the protein dimer interface or only one monomer. The free energy profile unequivocally points to the dimer mode as the thermodynamically favored one. The effect of these binding modes in stabilizing the protein dimer was also assessed, being related to its biological role in assisting the SARS viruses to bypass the host protective response. This work also constitutes a first step in the possible rational design of efficient therapeutic agents aiming at perturbing the interaction between SARS Unique Domain and guanine quadruplexes, hence enhancing the host defenses against the virus.



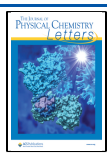
Coronaviruses inducing severe acute respiratory syndrome (SARS) gained first worldwide attention in 2003 following the outbreak of the SARS disease that mainly affected different southeastern Asian countries.[1−3] By the end of 2019 and the beginning of 2020, a novel coronavirus, styled SARS-CoV-2, had emerged and caused the COVID-19 pandemic.[4−7] At the time of writing this Letter, COVID-19 had spread in all continents, excluding Antarctica. As of April 5, 2020, more than 1 million confirmed cases had been reported, as well as more than 66 000 deaths.[8] Furthermore, severe limitations and social distancing measures had been implemented by an increasing number of countries, in an attempt to limit the spreading of the pandemic.[9] Indeed, coronaviruses, and SARS-CoV-2 in particular, are characterized by high infectious and transmissibility rates and may, in certain cases, severely attack the lungs and the respiratory system, an occurrence that can lead to the necessity of respiratory assistance and can even lead to death, especially in the presence of other comorbidities.[10−12]

Since the first outbreak of the pandemic, important scientific works appeared, aiming at understanding at a molecular level the mechanisms of action of SARS-CoV-2, and coronaviruses in general, that may be related to its transmissibility and mortality. Important successes have notably been achieved in both genome sequencing and structural resolution of the different viral proteins, also with the assistance of molecular modeling and simulation.[13−21] In fact, the possibility of finding efficient therapeutic strategies requires the elucidation of the fine mechanism that mediates the virus attack on the host cell, the resistance to the host immune system, and finally its virulence.[22−25]

Among the multifaced protein of coronaviruses, the so-called nonstructural protein 3 (Nsp3) represents a large protein, localized in the cytoplasm, consisting of about 2000 amino acids and comprising several different domains, also including a transmembrane region. Among the different domains of Nsp3,[26,27] whose precise functions have not been entirely clarified yet, the so-called SARS Unique Domain (SUD) deserves special attention, since it is present only in SARS-type coronaviruses and hence has been associated with the increased pathogenicity of this viral family.

The structure of SUD (presumably a common domain of different SARS viruses) has been resolved experimentally,[28−30] and it has been proved that the macrodomain is indeed constituted by a dimer of two symmetric monomers. Furthermore, both experimental and molecular docking

Received: April 8, 2020
Accepted: June 14, 2020
Published: June 15, 2020



https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

The Journal of Physical Chemistry Letters pubs.acs.org/JPCL Letter

investigations have pointed out a possible favorable interaction of SUD with nucleic acids, and in particular with RNA in G-quadruplex (G4) conformation.[28] The presence of a high density of lysine residues at the interface between two SUD monomers, forming a positively charged pocket, also suggests that RNA may be instrumental in favoring SUD dimerization, due to the negative charge of the RNA backbone hence suggesting the occurrence of electrostatic attraction. This observation may have a highly important biological implication since the dimerization has also been connected to the SUD native function. Tan et al.[28] have proposed that the ability of SUD to recognize and bind specific viral and/or host G4 sequences may have implications in regulating viral replication and/or hampering the host response to viral infection, as schematized in Figure 1. The hypothesis is based on the

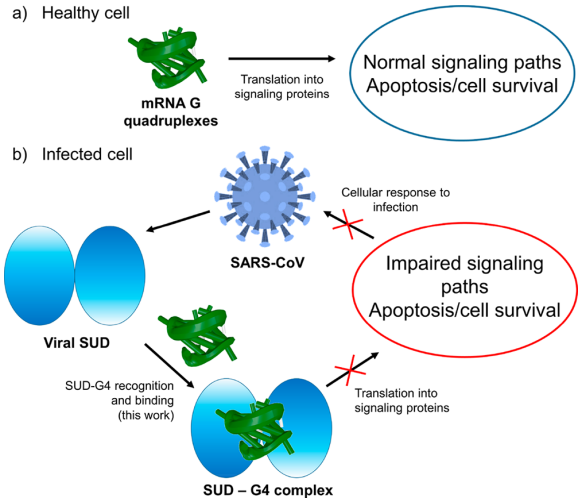


Figure 1. Schematic representation of the mRNA function in (a) a healthy cell and (b) an infected cell. Panel b shows the influence of viral SUD binding to G4 sequences of mRNA that encodes crucial proteins for the apoptosis/cell survival regulation and other signaling paths.

identification of G4 sequences in key host mRNA that encodes proteins involved in different signaling pathways such as apoptoting or survival signaling.[31−38] These proteins could induce a controlled cellular death of infected cells, thus slowing down or stopping the infection, or promote cell survival to delay apoptosis by producing antiviral cytokines.[37] However, the removal of the mRNA necessary to produce these signaling factors by viral SUD may impair the apoptosis/survival response pathways allowing massive cell infection.[28,37]

In this Letter, we report an extended all-atom molecular dynamics (MD) study of the interactions produced between a dimeric SUD domain and a short RNA G4 sequence. The crystal structure of the protein (pdb 2W2G) and of the oligonucleotide (pdb 1J8G)[39] have been chosen coherently with the experimental work performed by Tan et al.[28] Even though the chosen SUD starting structure belongs to the 2009 SARS-CoV, the very high nucleotide affinity[40] and the global conservation of the Nsp3 protein suggest that the RNA binding spots should be globally preserved. This is also further justified by the fact that SARS-CoV-2 Nsp3 has been recognized to suppress host gene expression and hence inhibit the immune response.[41] Equilibrium MD has allowed us to assess some of the hypothesized complexation modes between G4 and SUD, while also highlighting the most important interactions patterns at an atomistic level, and the effects of G4 in maintaining the dimer stability. Furthermore, to better sample the multidimensional conformational space and to quantify the strength of the interactions coming into play, the free-energy surface has been explored using enhanced sampling methods. A two-dimensional (2D) free energy profile has been computed along two coordinates defining the distance between the centers of mass of G4 and one SUD domain (G4-SUD_A), and the two SUD domains (SUD_A-SUD_B), respectively. The corresponding 2D potential of mean force (PMF) was obtained using a recently developed combination of extended adaptative biased force (eABF)[42] and metadynamics,[43] hereafter named meta-eABF.[44,45] Both protein and RNA have been described with the amber force field[46] including the bsc1 corrections,[47,48] and the MD simulations have been performed in the constant pressure and temperature ensemble (*NPT*) at 300 K and 1 atm. All MD simulations have been performed using the NAMD code[49] and analyzed via VMD,[50] the G4 structure has also been analyzed with the 3DNA suite.[51,52] More details on the simulation protocol can be found in the Supporting Information. To obtain starting conformations, the RNA was manually positioned in two different orientations close to the experimentally suggested SUD interaction area (see Supporting Information for the starting structure).[28] The equilibrium MD evolved, yielding two distinct interaction modes, as reported in Figure 2. In particular, we can easily distinguish between a first mode of binding in which the G4 mainly interacts with only one SUD monomer, called monomeric binding mode, and a second one in which the nucleic acid is firmly placed at the interface between the two protein monomeric subunits, referred to as the dimeric binding mode. Note that while for the monomeric mode we easily found a suitable starting point, two independent 200 ns MD trajectories were run to characterize the dimeric mode.

The corresponding root-mean-square deviations (RMSD) with respect to the initial structure are also reported and globally show that both the RNA and the protein units are stable. As expected, a slightly larger value of the RMSD is observed for the protein, as a consequence of its larger flexibility compared to the rigid G4 structures (Figure 2d,c). Note also that the slight initial increase of the protein RMSD observed for the dimeric mode is due to the necessity of a slight structural rearrangement to accommodate the G4 in the interaction pocket. Both modes are globally stable all along the MD trajectory, and no spontaneous release of the G4 is observed. At the end of the MD trajectory the dimeric mode exhibits a slight slipping toward the center of the interface area. This is probably due to the relative short length of the oligomer, also in agreement with previous experimental observations.[28]

Interestingly, the specific interaction patterns between G4 and the protein are different between the two binding modes, providing an extremely important different behavior. Indeed, the main driving force for the dimeric-like binding mode is the presence of extended electrostatic interactions between the negatively charged RNA backbone and the highly positive interaction pockets of the SUD complex. As a matter of fact, 11 Lys residues in the interaction pocket (shown in magenta in Figure 3a) allow persistent salt bridges and hydrogen bonds with the RNA phosphates. This finding is evidenced by the radial distribution function (RDF) between these positively

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

The Journal of Physical Chemistry Letters pubs.acs.org/JPCL Letter

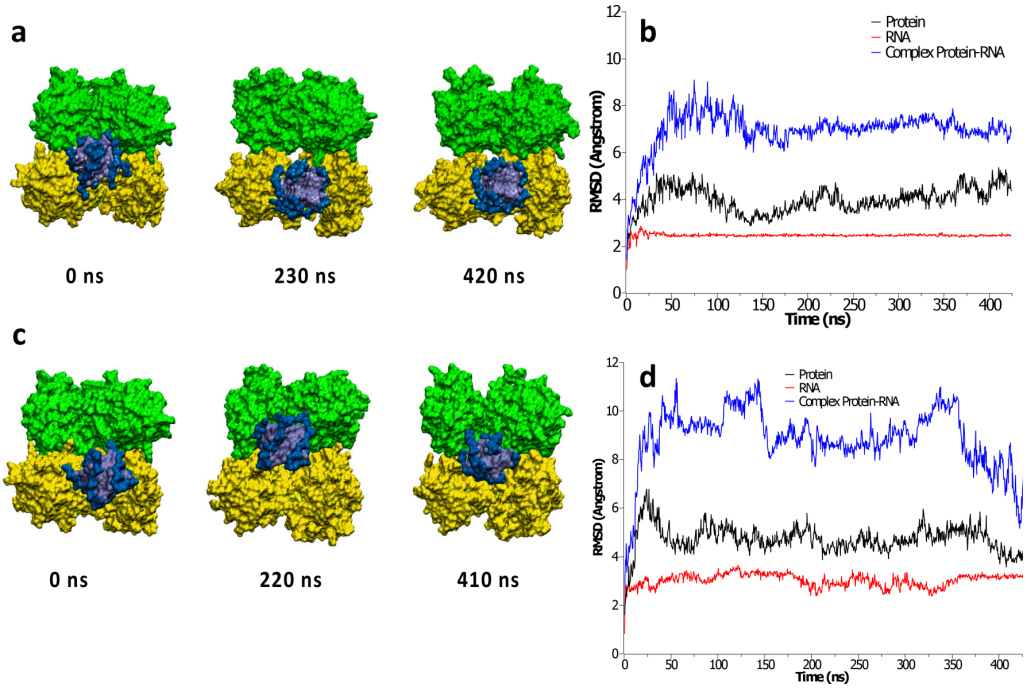


Figure 2. Representative snapshots of the SUD/G4 complex as extracted from the equilibrium MD simulation for the monomeric interaction mode (a) together with the corresponding RMSD for the protein and RNA fragments (b). Representative snapshots (c) and corresponding RMSD time evolutions (d) for the dimeric interaction mode. Note that the partial RMSD time series have been calculated aligning the system to the RNA or the protein, respectively.

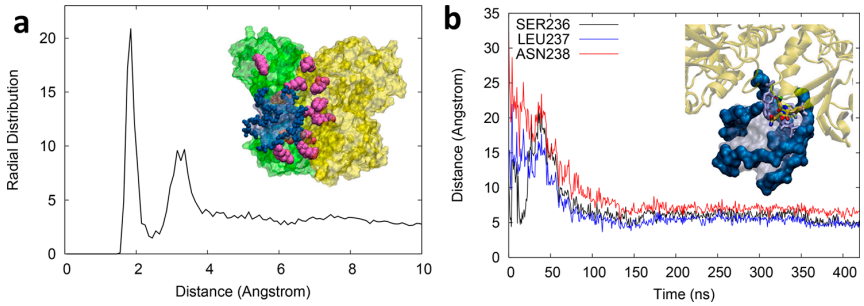


Figure 3. (a) RDF between the RNA phosphate oxygens and the Lys NH_3 hydrogens for the dimeric interaction mode (see Figure 2c). The inlay shows a representative snapshot showing the network of Lys (in purple, van der Waals representation) in the interaction pocket surrounding the RNA G4, whose backbone phosphate groups are highlighted in dark blue. (b) Time series of the distances between the α-carbon of Ser236, Leu237, and Asn238 and the G4 nearby uracil or guanine oxygen atom. A zoom of the representative snapshot (Figure 2a) showing the corresponding interactions is also provided.

charged lysine side chains and the negatively charged phosphate oxygen atoms of G4 (depicted in dark blue in Figure 3b), which shows a very intense and sharp peak at around 2 Å (Figure 3a). Interestingly, a secondary peak in the RDF is also observed at 3.5 Å, probably defining a second layer of interaction patterns that should contribute to the overall stabilization of the binding. Conversely, the monomeric mode is driven by interactions mainly involving the terminal uracil moieties and the top guanine leaflet instead of the phosphate backbone of G4. As shown in Figure 3b, hydrophobic interactions emerge through the action of a triad of amino acids, namely, Ser236, Leu237, and Asn238, that position themselves on top of the terminal leaflet also interacting with the peripheral uracil nucleobases, in a mode strongly resembling the top-binding experienced by a number of G4 drugs.[53−55] This is nicely confirmed by the analysis of the time series of the distance between the α-carbon of these amino acids and the nearby guanine that readily drops at around 5 Å and stays remarkably stable all along the MD. Interestingly, the interaction is sufficiently strong to induce a partial deformation of the planarity of the G4 leaflets. Even though from considerations based on chemical intuition those interactions could be referred to as mainly due to dispersion, the inherent parametrization of the amber force field does not allow us to completely disentangle and decompose the polarization, dispersion, and electrostatic contributions. The fact that the monomeric binding mode is driven by noncovalent interactions with one of the exposed G4 leaflets may also contribute

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

explaining the fact that longer G4 sequences are preferentially recognized by the SUD interface region. Indeed, in this case, for obvious statistic reasons, the ratio between the interaction with the backbone or with the terminal leaflet clearly favors the former. However, the monomeric mode may also act efficiently in the process of recruitment of RNA, either viral or cellular, efficiently anchoring the oligomer that can subsequently be safely displaced through the interface area.

Apart from the different positioning of the G4, other structural evidences can already be surmised from the visual inspection of the MD trajectory. In particular, the SUD dimers appear more compact and the interface region better conserved when the RNA G4 adopts the dimeric binding mode, as can also been appreciated in Figure 4. These results clearly indicate that the dimeric mode leads to a greater stability of the G4-SUD complex.

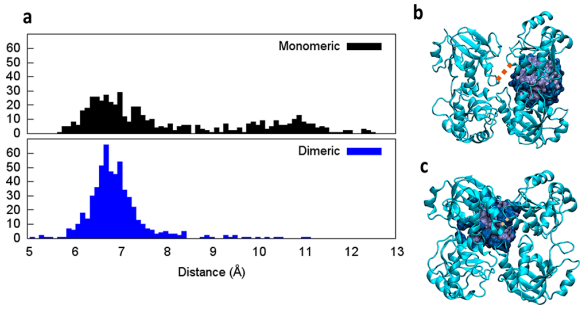


Figure 4. Distribution of the distances between the α-carbon of Arg266 and Ala366 (a) forming a tweezer in the SUD subdomains interface region and representative snapshots showing this region for the monomer (b) and the dimer (c) conformation. The dashed orange line indicates the position of the protein residues corresponding to the distance measured.

Indeed, the presence of G4 in the dimeric binding mode induces rigidity in the SUD subunits interface due to the formation of the complex. This effect can be quantified by analyzing the distribution of the distance of a tweezer in the SUD interface area formed by the residues Ala355 and Arg266, as reported in Figure 4. Indeed, while in the case of the dimeric-like conformation a peaked distribution centered at relative close distances (6.7 Å) is observed, indicating a closed and quite rigid disposition, a much broader and bimodal distribution is found for the monomer-like conformation, presenting, most notably, a secondary maximum at about 11 Å, which confirms the partial destabilization of the SUD subdomains interface and the greater flexibility induced by this binding mode.

To further examine the conformational space spanned by the G4/SUD complex, and in particular the role of the RNA in favoring the dimerization and the structure of the interface, we resorted to enhanced sampling MD simulations to obtain the 2D free energy profile along two relevant collective variables: first, the distance between G4 and SUD, and second, the separation between the two SUD subdomains. Our choice of collective variables does not allow us to explore the binding between the two surfaces of the SUD domain; however, from the results of Tan et al.[28] it is clear that the interaction with RNA takes place preferentially at the positively charged interface. However, our methodology is perfectly adapted to characterize the influence of the binding of RNA to the stabilization of the interface between the two SUD monomers since it allows the sliding of the G4 on the SUD surface. The PMF is reported in Figure 5 together with representative

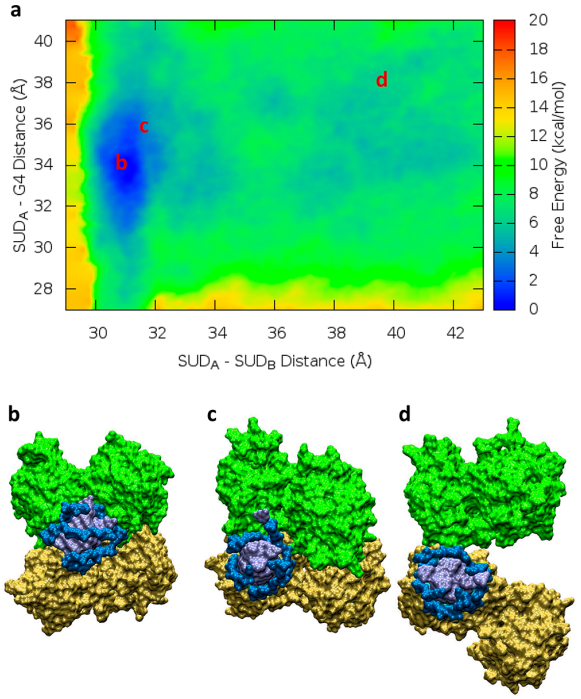


Figure 5. (a) 2D free energy profile describing the interaction with RNA and SUD dimerization. Representative snapshots are also provided describing the principal minimum (b), the secondary minimum (c), and an open SUD conformation (d). The position of the selected snapshots on the PMF map is also indicated in red.

snapshots along the reaction coordinates. From the analysis of the PMF, one can evidence the presence of an evident minimum in the free energy profile corresponding to the situation in which the G4 is interacting through the dimer mode, in which the SUD dimer is compact (Figure 5b). The free-energy stabilization, with respect to the situation in which the G4, is well separated from the protein amounts to about 6 kcal/mol. Around the principal minimum a slightly less stable and extended region is also observed having a stabilization free energy of about 3−4 kcal/mol and corresponding to the sliding of G4 in the monomer conformation (Figure 5c). The rest of the free energy surface appears instead rather flat, and no appreciable barriers are observed along the collective variable. The topology of the free energy surface hence accounts for the possibility to observe important conformational movements, leading to open conformations in which the SUD subdomain interface has been basically destroyed (Figure 5d). However, such conformations are instead hampered by the dimer-like conformation of the RNA. The free energy map unambiguously shows that the dimer mode is the preferred one and also confirms the role of the G4 binding in maintaining the dimeric SUD conformation, since no appreciable free energy barrier for the opening of the SUD dimer is observed when the RNA is unbound. Thus, the dimer mode binding site clearly constitutes a specific target that may help in the development of new efficient antiviral agents against coronaviruses.

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

By using a combination of equilibrium and enhanced sampling MD, we have identified the structure and dynamics of the viral SUD dimers and their interaction with RNA oligomers forming guanine quadruplexes. Even if the accuracy of the obtained results is obviously dependent on the choice of the specific force field and of the specific collective variables, we believe we have been able to offer a significant sampling of the binding between SUD and RNA G4, and of its effects on the stabilization of the SUD complex and the monomer−monomer interface. In particular, we have clearly evidenced the existence of two different interaction modes, one happening at the subdomains interface (dimeric binding mode) and another involving only one of the SUD monomers. While the former mode is mostly driven by electrostatic interactions and appears to be more stable and is essential in rigidifying the protein dimer and stabilizing the SUD-G4 interface, the latter assures a great flexibility to the protein and is guided by noncovalent interactions. Such stable interaction between SUD and G4-RNA can be related to the translational default of the concerned proteins (see Figure 1). Furthermore, the monomer interaction mode, happening in a more solvent exposed region of the protein, can also play a role in recruiting RNA fragments in the first step of the SUD-G4 recognition.

The SUD dimer is thought to have an important biological role in allowing coronaviruses to bypass host cell protective and immune systems, hence allowing the virus morbidity and transmissibility. Indeed, and even if in the present work we used a protein belonging to SARS-CoV, we believe that the mechanism underlined is quite general and may represent one of the reasons of the high pathogenicity of SARS agents. As a matter of fact, the role of SARS-CoV-2 Nsp3 in suppressing host gene expression has also been recently evidenced.[41] The destabilization of the SUD/RNA complex, which could be achieved for instance by specific G4 ligands, or by competitors for the two G4 recognition spots evidenced in SUD, may constitute appealing potential therapeutic strategies. This is particularly true since these protein domains are shared by the family of SARS-type viruses and thus represent a molecular target that could stay operative in the case of future mutations, leading to new viral strains. As a continuation of the present work, we plan to address the study of the interaction between specific ligands and RNA G4 oligomers, also taking into account the perturbation induced on the recognition by SUD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c01097.

Details of the protocol used for equilibrium MD and free energy determination, analysis of the structural parameters of the G4 unit and of the setup of the SUD initial structure, time series of the distance between the residues defining the opening of SUD (PDF)

Starting structure of the simulations (PDB)

Starting structure of the simulations (PDB)

■ AUTHOR INFORMATION

Corresponding Authors

Marco Marazzi − *Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering and Chemical Research Institute "Andrés M. del Río" (IQAR), Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain;* orcid.org/0000-0001-7158-7994; Email: marco.marazzi@uah.es

Giampaolo Barone − *Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, Università degli Studi di Palermo, 90128 Palermo, Italy;* orcid.org/0000-0001-8773-2359; Email: giampaolo.barone@unipa.it

Antonio Monari − *Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France;* orcid.org/0000-0001-9464-1463; Email: antonio.monari@univ-lorraine.fr

Authors

Cécilia Hognon − *Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France*

Tom Miclot − *Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France; Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, Università degli Studi di Palermo, 90128 Palermo, Italy*

Cristina García-Iriepa − *Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering and Chemical Research Institute "Andrés M. del Río" (IQAR), Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain;* orcid.org/0000-0002-7577-8242

Antonio Francés-Monerris − *Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France; Departament de Química Física, Universitat de València, 46100 Burjassot, Spain;* orcid.org/0000-0001-8232-4989

Stephanie Grandemange − *Université de Lorraine and CNRS, CRAN UMR 7039, F-54000 Nancy, France*

Alessio Terenzi − *Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, Università degli Studi di Palermo, 90128 Palermo, Italy;* orcid.org/0000-0001-9751-1373

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.0c01097

Author Contributions

▽C.H. and T.M. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support from the Université de Lorraine and CNRS is gratefully acknowledged. T.M. thanks University of Palermo for funding a joint Ph.D. program A.F.-M. is grateful to Generalitat Valenciana and the European Social Fund for a postdoctoral contract (APOSTD/2019/149) and the *Ministerio de Ciencia e Innovación* (project (CTQ2017-87054-C2-2-P) for financial support. French CNRS and IDRIS national computing center are acknowledged for graciously providing access to computational resources in the framework of the special COVID-19 mobilization under the project "Seek&-Destroy". Part of the calculations have been performed on the LPCT local computing cluster and on the regional ExpLor center in the frame of the project "Dancing Under the Light".

■ REFERENCES

(1) Xing, W.; Hejblum, G.; Leung, G. M.; Valleron, A. J. Anatomy of the Epidemiological Literature on the 2003 SARS Outbreaks in Hong Kong and Toronto: A Time-Stratified Review. *PLoS Medicine.* **2010**, 7, e1000272.

(2) Stockman, L. J.; Bellamy, R.; Garner, P. SARS: Systematic Review of Treatment Effects. *PLoS Medicine.* **2006**, 3, e343.

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

(3) Ahmad, A.; Krumkamp, R.; Reintjes, R. Controlling SARS: A Review on China's Response Compared with Other SARS-Affected Countries. *Trop. Med. Int. Health* **2009**, *14*, 36−45.
(4) Watkins, J. Preventing a Covid-19 Pandemic. *BMJ.* **2020**, m810.
(5) Tang, B.; Wang, X.; Li, Q.; Bragazzi, N. L.; Tang, S.; Xiao, Y.; Wu, J. Estimation of the Transmission Risk of 2019-NCov and Its Implication for Public Health Interventions. *SSRN Electron. J.* **2020**, DOI: 10.2139/ssrn.3525558.
(6) Cao, Z.; Zhang, Q.; Lu, X.; Pfeiffer, D.; Jia, Z.; Song, H.; Zeng, D. D. Estimating the Effective Reproduction Number of the 2019-NCoV in China. *medRxiv* **2020**, DOI: 10.1101/2020.01.27.20018952.
(7) Bedford, J.; Enria, D.; Giesecke, J.; Heymann, D. L.; Ihekweazu, C.; Kobinger, G.; Lane, H. C.; Memish, Z.; Oh, M. don; Sall, A. A. COVID-19: Towards Controlling of a Pandemic. *Lancet* **2020**, *395*, 1015−1018.
(8) WHO. Coronavirus (COVID-19) Events as They Happen. 2020; https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
(9) Hellewell, J.; Abbott, S.; Gimma, A.; Bosse, N. I.; Jarvis, C. I.; Russell, T. W.; Munday, J. D.; Kucharski, A. J.; Edmunds, W. J.; Sun, F.; et al. Feasibility of Controlling COVID-19 Outbreaks by Isolation of Cases and Contacts. *Lancet Glob. Heal.* **2020**, *8* (4), e488−e496.
(10) Verity, R.; Okell, L. C.; Dorigatti, I.; Winskill, P.; Whittaker, C.; Imai, N.; Cuomo-dannenburg, G.; Thompson, H.; Walker, P. G. T.; Fu, H.; et al. Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis. *Lancet Infect. Dis.* **2020**, *20*, 669−677.
(11) Vincent, J.-L.; Taccone, F. S. Understanding Pathways to Death in Patients with COVID-19. *Lancet Respir. Med.* **2020**, *8*, 430.
(12) Yuan, J.; Li, M.; Lv, G.; Lu, Z. K. Monitoring Transmissibility and Mortality of COVID-19 in Europe. *Int. J. Infect. Dis.* **2020**, *95*, 311.
(13) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science (Washington, DC, U. S.)* **2020**, *367* (6483), 1260−1263.
(14) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural Basis for the Recognition of the SARS-CoV-2 by Full-Length Human ACE2. *Science (Washington, DC, U. S.)* **2020**, *367* (6485), 1444.
(15) Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; et al. Analysis of Therapeutic Targets for SARS-CoV-2 and Discovery of Potential Drugs by Computational Methods. *Acta Pharm. Sin. B* **2020**, *10*, 766.
(16) Wang, C.; Li, W.; Drabek, D.; Okba, N. M. A.; van Haperen, R.; Osterhaus, A. D. M. E.; van Kuppeveld, F. J. M.; Haagmans, B. L.; Grosveld, F.; Bosch, B.-J. A Human Monoclonal 1 Antibody Blocking SARS-CoV-2 Infection. *Nat. Commun.* **2020**, *11*, 3964.
(17) Joyce, M. G.; Sankhala, R. S.; Chen, W.-H.; Choe, M.; Bai, H.; Hajduczki, A.; Yan, L.; Sterling, S. L.; Peterson, C.; Green, E. C.; et al. A Cryptic Site of Vulnerability on the Receptor Binding Domain of the SARS-CoV-2 Spike Glycoprotein. *bioRxiv* **2020**, DOI: 10.1101/2020.03.15.992883.
(18) Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; et al. Inhibition of SARS-CoV-2 (Previously 2019-NCoV) Infection by a Highly Potent Pan-Coronavirus Fusion Inhibitor Targeting Its Spike Protein That Harbors a High Capacity to Mediate Membrane Fusion. *Cell Res.* **2020**, *30* (4), 343−355.
(19) Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; et al. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11*, 1620.
(20) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281.
(21) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science (Washington, DC, U. S.)* **2020**, eabb3405.
(22) Rismanbaf, A. Potential Treatments for COVID-19; a Narrative Literature Review. *Arch. Acad. Emerg. Med.* **2020**, *8* (1), e29.
(23) Beck, B. R.; Shin, B.; Choi, Y.; Park, S.; Kang, K. Predicting Commercially Available Antiviral Drugs That May Act on the Novel Coronavirus (2019-NCoV), Wuhan, China through a Drug-Target Interaction Deep Learning Model. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 784.
(24) Mitjà, O.; Clotet, B. Use of Antiviral Drugs to Reduce COVID-19 Transmission. *The Lancet Global Health* **2020**, DOI: 10.1016/S2214-109X(20)30114-5.
(25) Bzowka, M.; Mitusinska, K.; Raczynska, A.; Samol, A.; Tuszynski, J. A.; Gora, A. Molecular Dynamics Simulations Indicate the COVID-19 Mpro Is Not a Viable Target for Small-Molecule Inhibitors Design. *bioRxiv* **2020**, DOI: 10.1101/2020.02.27.968008.
(26) Phillips, J. E.; Jackwood, M. W.; McKinley, E. T.; Thor, S. W.; Hilt, D. A.; Acevedol, N. D.; Williams, S. M.; Kissinger, J. C.; Paterson, A. H.; Robertson, J. S.; et al. Changes in Nonstructural Protein 3 Are Associated with Attenuation in Avian Coronavirus Infectious Bronchitis Virus. *Virus Genes* **2012**, *44* (1), 63−74.
(27) Graham, R. L.; Sparks, J. S.; Eckerle, L. D.; Sims, A. C.; Denison, M. R. SARS Coronavirus Replicase Proteins in Pathogenesis. *Virus Res.* **2008**, *133* (1), 88−100.
(28) Tan, J.; Vonrhein, C.; Smart, O. S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J. R.; Schmidt, C. L.; Hilgenfeld, R. The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes. *PLoS Pathog.* **2009**, *5* (5), No. e1000428.
(29) Johnson, M. A.; Chatterjee, A.; Neuman, B. W.; Wüthrich, K. SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding. *J. Mol. Biol.* **2010**, *400* (4), 724−742.
(30) Chatterjee, A.; Johnson, M. A.; Serrano, P.; Pedrini, B.; Joseph, J. S.; Neuman, B. W.; Saikatendu, K.; Buchmeier, M. J.; Kuhn, P.; Wuthrich, K. Nuclear Magnetic Resonance Structure Shows That the Severe Acute Respiratory Syndrome Coronavirus-Unique Domain Contains a Macrodomain Fold. *J. Virol.* **2009**, *83* (4), 1823−1836.
(31) Fehr, A. R.; Perlman, S. In *Coronaviruses: An Overview of Their Replication and Pathogenesis BT - Coronaviruses: Methods and Protocols*; Maier, H. J., Bickerton, E., Britton, P., Eds.; Springer New York: New York, NY, 2015; pp 1−23.
(32) Brázda, V.; Fojta, M. The Rich World of P53 DNA Binding Targets: The Role of DNA Structure. *Int. J. Mol. Sci.* **2019**, *20* (22), 5605.
(33) Lei, J.; Kusov, Y.; Hilgenfeld, R. Nsp3 of Coronaviruses: Structures and Functions of a Large Multi-Domain Protein. *Antiviral Res.* **2018**, *149*, 58−74.
(34) Kusov, Y.; Tan, J.; Alvarez, E.; Enjuanes, L.; Hilgenfeld, R. A G-Quadruplex-Binding Macrodomain within the "SARS-Unique Domain" Is Essential for the Activity of the SARS-Coronavirus Replication−Transcription Complex. *Virology* **2015**, *484*, 313−322.
(35) Wei, J. S.; Whiteford, C. C.; Cenacchi, N.; Son, C. G.; Khan, J. BBC3Mediates Fenretinide-Induced Cell Death in Neuroblastoma. *Oncogene* **2005**, *24* (54), 7976−7983.
(36) Garcia-Saez, I.; Tcherniuk, S.; Kozielski, F. The Structure of Human Neuronal Rab6B in the Active and Inactive Form. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62* (7), 725−733.
(37) Mizutani, T. Signal Transduction in SARS-CoV-Infected Cells. *Ann. N. Y. Acad. Sci.* **2007**, *1102* (1), 86−95.
(38) Jin, G.; Klika, A.; Callahan, M.; Faga, B.; Danzig, J.; Jiang, Z.; Li, X.; Stark, G. R.; Harrington, J.; Sherf, B. Identification of a Human NF-KB-Activating Protein, TAB3. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (7), 2028−2033.
(39) Deng, J.; Xiong, Y.; Sundaralingam, M. X-Ray Analysis of an RNA Tetraplex (UGGGGU)4 with Divalent Sr2+ Ions at Subatomic Resolution (0.61 Å). *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (24), 13665−13670.
(40) Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. Network-Based Drug Repurposing for Novel Coronavirus 2019-NCoV/SARS-CoV-2. *Cell Discovery* **2020**, *6*, 14.

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

(41) Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M.; Kratzat, H.; Hayn, M.; Mackens-Kiani, T.; Cheng, J.; et al. Structural Basis for Translational Shutdown and Immune Evasion by the Nsp1 Protein of SARS-CoV-2. *bioRxiv* **2020**, DOI: 10.1101/2020.05.18.102467.
(42) Zhao, T.; Fu, H.; Lelièvre, T.; Shao, X.; Chipot, C.; Cai, W. The Extended Generalized Adaptive Biasing Force Algorithm for Multidimensional Free-Energy Calculations. *J. Chem. Theory Comput.* **2017**, *13* (4), 1566−1576.
(43) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826−843.
(44) Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *J. Phys. Chem. Lett.* **2018**, *9* (16), 4738−4745.
(45) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **2019**, *52* (11), 3254−3264.
(46) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.
(47) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114−4127.
(48) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11* (12), 5723−5736.
(49) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(50) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(51) Lu, X. J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108−5121.
(52) Lu, X. J.; Olson, W. K. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3* (7), 1213−1227.
(53) Terenzi, A.; Gattuso, H.; Spinello, A.; Keppler, B. K.; Chipot, C.; Dehez, F.; Barone, G.; Monari, A. Targeting G-Quadruplexes with Organic Dyes: Chelerythrine−DNA Binding Elucidated by Combining Molecular Modeling and Optical Spectroscopy. *Antioxidants* **2019**, *8* (10), 472.
(54) Lauria, A.; Bonsignore, R.; Terenzi, A.; Spinello, A.; Giannici, F.; Longo, A.; Almerico, A. M.; Barone, G. Nickel(II), Copper(II) and Zinc(II) Metallo-Intercalators: Structural Details of the DNA-Binding by a Combined Experimental and Computational Investigation. *Dalton Trans.* **2014**, *43* (16), 6108−6119.
(55) Terenzi, A.; Bonsignore, R.; Spinello, A.; Gentile, C.; Martorana, A.; Ducani, C.; Högberg, B.; Almerico, A. M.; Lauria, A.; Barone, G. Selective G-Quadruplex Stabilizers: Schiff-Base Metal Complexes with Anticancer Activity. *RSC Adv.* **2014**, *4* (63), 33245−33256.

https://dx.doi.org/10.1021/acs.jpclett.0c01097
J. Phys. Chem. Lett. 2020, 11, 5661−5667

CHAPTER 15

Specific recognition of *c-Myc* by DARPin 2E4

Several strategies have been developed in order to identify and localize G-quadruplexes *in cellulo*, for analytical or therapeutic purposes. Generally, these involve small molecules that interact with G-quadruplexes [200, 201, 202]. Also, antibodies have been designed for the same purpose [203], however, the design of antibodies is often complex. Another class of protein, named DARPins, has been identified as being able to interact preferentially with G-quadruplexes [204]. DARPins are a class of synthetic protein inspired by natural ankyrin proteins. More precisely, they are small repeating proteins consisting of a succession of ankyrin motifs: two helices and a loop. Each motif is called a module. DARPins have an N-terminal module and a C-terminal module, called caps. Between the caps, there are one or more internal modules that are designed according to the purpose [205]. DARPins have found multiple applications as cell marker, protein cap to facilitate crystalography or as therapeutic agent [206, 205, 207]. In 2014 Scholz, Hansen, and Plückthun [204] performed a study highlighting the recognition specificity of DARPin 2E4 towards the G-quadruplex of the *c-Myc* promoter. The specificity is extremely well defined experimentally, although the exact mechanism of recognition is still not known. Accordingly, the following paper presents an approach to provide an explanation of the mechanisms behind the specific recognition. Our strategy is based on homology modeling, docking and molecular dynamics simulations. By comparing the behavior of DARPin 2E4, specific to *c-Myc*, and that of DARPin 2G10, non-specific, towards three G-quadruplexes: *c-Myc*, *h-Telo* and *Bcl-2*. Analysis shows that the specificity of the 2E4/*c-Myc* interaction arises from the recognition of a structural motif induced by a peculiar folding of the *c-Myc* G-quadruplex.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

G-Quadruplex Recognition by DARPIns through Epitope/Paratope Analogy**

Tom Miclot,*[a, b] Emmanuelle Bignon,[b] Alessio Terenzi,[a] Stéphanie Grandemange,[c] Giampaolo Barone,*[a] and Antonio Monari*[b, d]

Abstract: We investigated the mechanisms leading to the specific recognition of Guanine Guadruplex (G4) by DARPins peptides, which can lead to the design of G4 s specific sensors. To this end we carried out all-atom molecular dynamic simulations to unravel the interactions between specific nucleic acids, including human-telomeric (h-telo), Bcl-2, and *c-Myc*, with different peptides, forming a DARPin/G4 complex. By comparing the sequences of DARPin with that of a peptide known for its high affinity for *c-Myc*, we show that the recognition cannot be ascribed to sequence similarity but, instead, depends on the complementarity between the three-dimensional arrangement of the molecular fragments involved: the α-helix/loops domain of DARPin and the G4 backbone. Our results reveal that DARPins tertiary structure presents a charged hollow region in which G4 can be hosted, thus the more complementary the structural shapes, the more stable the interaction.

Introduction

In addition to the well-known double helical arrangement, the important biological role of non-canonical nucleic acid is nowadays widely recognized. Among the different non-canonical DNA or RNA structures, guanine quadruplexes (G4 s) are highly studied and characterized.[1–3] From a chemical point of view, G4 s are formed in guanine-rich nucleic acids, whose nucleobases develop primarily cooperative Hoogsteen-type hydrogen bonds. Because of the specificity of this interaction, DNA (or RNA) is then organized in a series of stacked quartets, whose macromolecular arrangement is further stabilized by a metal cation occupying the central channel. Three main topologies can be adopted by non-canonical G4 structures depending on the 5′-3′ orientation of the strands forming the G4 s backbone: parallel, antiparallel and hybrid (Figure 1). In the parallel conformation, all the strands are oriented in the same way, in the antiparallel conformation the strands are inversely oriented two by two, while in the hybrid conformation only one strand is inversely oriented with respect to the other three.[4]

G4 s have been identified in cellular nucleic acids and have been associated to the control of key biological functions. As a matter of fact, G4 s are involved in gene regulations,[5,6] in

[a] *T. Miclot, Prof. A. Terenzi, Prof. G. Barone*
Department of Biological, Chemical and Pharmaceutical Sciences and Technologies
Università degli Studi di Palermo
Viale delle Scienze, 90128 Palermo (Italy)
E-mail: tom.miclot@unipa.it
giampaolo.barone@unipa.it
[b] *T. Miclot, Dr. E. Bignon, Prof. A. Monari*
Université de Lorraine and CNRS LPCT UMR 7019
54000 Nancy (France)
[c] *Prof. S. Grandemange*
Université de Lorraine and CNRS CRAN UMR 7039
54000 Nancy (France)
[d] *Prof. A. Monari*
Université Paris Cité and CNRS, ITODYS
75006 Paris (France)
E-mail: antonio.monari@u-paris.fr

[**] *A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.1101/2022.06.13.495947).*
Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202201824

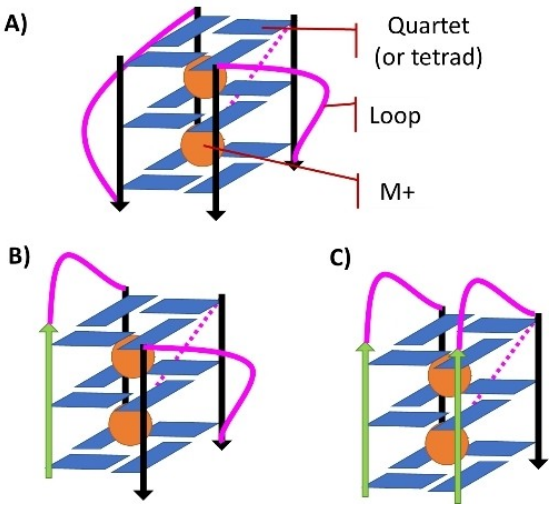


Figure 1. Schematic representation of the different topologies of G4. A) The parallel form in which all strands are parallel to each other, B) the hybrid form in which one strand is oriented antiparallel to all the others, and C) the antiparallel form in which the strands are antiparallel two by two.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

neurodegenerative diseases,[7,8] in the induction of DNA damages and in oncogenesis.[9–12] Furthermore, they have been recognized to play a role in the regulation of cellular cycles and in the regulation of post-translational modification in proteins.[12–15] G4 s have also been identified in both DNA and RNA viral genomes, including SARS-CoV-2,[16–18] where they may exert vital functions in regulating viral infection cycles.[19–22] Obviously, all these processes can only take place through a molecular machinery involving proteins selectively recognizing specific DNA or RNA G4 s. Among them we may cite ATP-dependent DNA/RNA helicase DHX36,[23] G-rich sequence factor 1,[24] or fragile X mental retardation protein (SFMRP).[25] Thus, the development of artificial or biomimetic specific G4 binders, capable to recognize either DNA or RNA, is highly valuable. Furthermore, such ligands may be exploited either in a therapeutic context or for the rapid identification and localization of G4 s in cells or cellular compartments.[26–29] Protein engineering has also led to the development of antibodies presenting specificity and selectivity toward G4 s. In this case, the recognition of G4 s proceeds through the epitope/paratope mechanism, in which the G4 acts as the epitope of an antibody.[30] However, the design of antibodies is definitively not straightforward, and their use is typically limited to the identification of the subcellular localization of G4 s.[30–33] Smaller peptides specifically recognizing G4 s and even discriminating between different G4 types have also been proposed. This is typically the case of DARPins,[34] a class of synthetic proteins derived from the modification of natural ankyrins and mostly known as chaperone agents in crystallography.[35] In addition, DARPins have also been used as cellular markers in biological imaging and for therapeutic purposes.[36,37]

Understanding the factors underlying the specific recognition of G4 s by DARPins can facilitate the design of sensors able to discriminate the G4 s subcellular localization and their specific sequences. In this contribution, we model the DARPin/G4 interaction and, thus, unravel the specific recognition modes by combining molecular docking and long-scale all atom molecular dynamics (MD) simulations. We focus on 2E4 DARPin, which is specific for the G4 present in the *c-Myc* oncogene promoter, and 2G10, which has a slight specificity for different G4 s.[34] As for the nucleic acid counterpart, we restrict our study to the human telomeric G4, as well as the G4 s in the *c-Myc*[34] and *Bcl-2* promoters.

Results and Discussion

MD simulations shed light on the structural details underlying the specific DARPins/G4 interaction. Indeed, by sampling the conformational space through different initial interaction positions, it is possible to analyze whether the interaction is conserved, the binding of the G4 affects the flexibility, the nucleic acid rearranges to reach a more stable pose, or if the proposed DARPin/DNA complex is not stable and separates. In our case most of the G4 s/DARPin complexes are persistent and stable all along the MD and the peptides interact with the G4 s through regions composed of large loops and helices, which overlap well with the recognized canonical interaction zones of the DARPins.

The only exceptions can be highlighted for 2E4/*c-Myc* which in one of the poses leads to a very labile and mobile interaction between G4 and the protein as confirmed by clustering yielding two dominant structures, representing 41.71% and 30.56% of the trajectory, respectively. On the other hand, 2E4/*h-Telo* (64.33% of the trajectory) and 2G10/*c-Myc* (75.10%) yield dominant clusters that interacts only through the loop ends of DARPins and one or two nucleotides of the flexible G4 s loops. (All the clustered structures can be found in the Supporting Information).

Residue-scale analysis of the G4/DARPin complexes

Before exposing the structural details of the G4 s/DARPin complexes at the atomistic scale, it is interesting to consider the interaction at a residue-level scale, and in particular classify the different interaction patterns in terms of the number of involved nucleic acid or protein residues. Figure 2 shows all the residues which remains within a cutoff of 3 Å from either the protein or the nucleic acid with a frequency at least equal to 50% of the simulation time. On average six nucleic acid residues of *h-Telo* and eight amino acids of 2G10 can be identified. However, 2G10 interacts persistently through only five amino acids with *c-Myc* and *Bcl-2* which in turn only bring a maximum of two or three nucleotides into persistent contact with the protein. Conversely, the 2E4/*h-Telo* interaction appears to be driven by three nucleotides and six amino acids. For 2E4, the interaction gathers eight amino acids with both *c-Myc* and *Bcl-2*, yet a different number of nucleic acid residues is involved, that is, four for *Bcl-2* and seven for *c-Myc*.

This first analysis, at the residue level, already draws a general picture of the specific recognition of G4 by DARPins, confirming that the protein/nucleic acid recognition is favored by a high number of interacting residues. However, it needs to be completed identifying the exact nature of the interacting residues, their specific frequency, and the specific structural features.

Three different modes of interaction leading to the G4/DARPin complexes

To improve the global scale analysis presented in the previous subsection we should identify a region of the G4 s that is selectively recognized by DARPins. As a matter of fact, *h-Telo* does not show any specific interacting region or hotspot with 2E4 or 2G10 (see Figure 3A). This could confirm that the non-specific recognition of *h-Telo* is due to the absence of a well-defined target region on this nucleotide. In contrast, more pronounced specific interaction regions may be recognized in the two others G4 s. Indeed, it can be seen in Figure 3(B) that two interaction areas clearly stand out for *Bcl-2* interacting with either 2EA or 2G10, that is, the one including residues dC5 to dG8 and the one involving residues dG20 to dG22. Since the

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

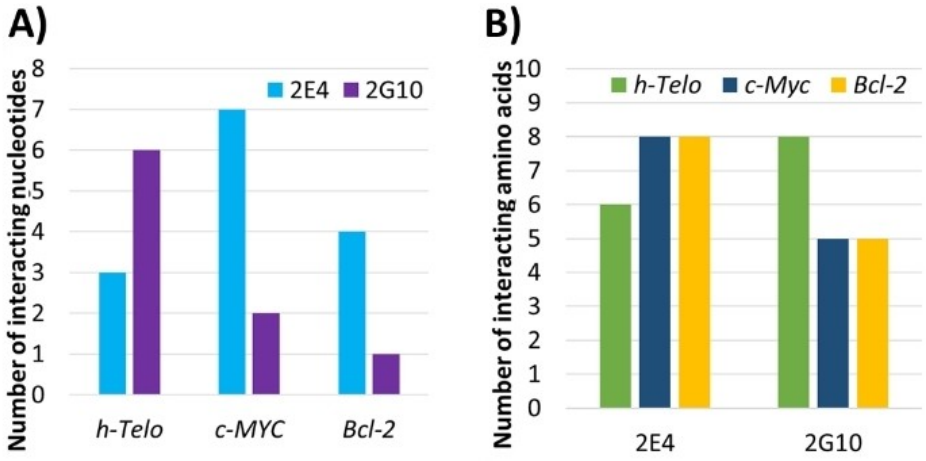


Figure 2. Number of A) nucleic acids and B) amino acids residues involved in the protein/DNA interaction and found at a frequency of at least 50% of the simulation time.

same nucleic acid regions are evidenced for both DARPins, the interaction mode can be classified as structurally similar in each complex. Thus, no specific recognition of *Bcl-2* by 2E4 or 2G10 can be inferred, since such specific recognition should involve interaction areas that must differ between two different DARPins. This is, indeed, the case for the G4 present in the *c-Myc* promoter. Figure 3(C) clearly shows regions of very pronounced contact and different for each of the DARPins. For the interaction with 2G10, the hot spot includes residues dG15 to dT20, although the contact frequencies are still quite spread across the whole G4. On the other hand, 2E4 highlights two very strong and localized contact points. The most important one concerns residues dG6 to dA12, while the second one corresponds to the last two residues of G4, dA21 and dA22. Thus, the specific recognition of *c-Myc* by 2E4 could be achieved either through the recognition of its sequence, or through a specific structural motif. Our analysis indicates three possible scenarios: 1) a rather general interaction that does not involve any specific G4 region or sequence (*h-telo*); 2) a non-specific interaction involving particular G4 s regions, which are however recognized by all the proteins (*Bcl-2*); 3) a specific interaction driven by a few nucleotides having very high contact frequencies with specific DARPin (*c-Myc*).

Identification of a putative selective DARPin interaction area

Repeating the same analysis while focusing on the protein counterpart we identify the amino acids mostly involved in the recognition of the non-canonical DNA structure. Similarly, to what has already been observed for G4 s, selectivity should correlate with few specific amino acids having high contact frequencies with G4 s. Conversely, for non-selective recognition a more scattered distribution of the interaction frequencies should be observed.

The distribution of the interaction contacts of 2G10 with the three G4 s (Figure 4A) shows three distinct peaks. The first one corresponds to residues N34 and I35, the second one gathers residues R67, W68, R70, K78 and W79, while the last one comprises residues K100 and K101. While these localized protein areas certainly correspond to a strong interaction with G4, they appear rather non-specific since they are present for all the three G4 s. However, the interaction with *h-Telo* is also driven by amino acids whose contact frequency was low or zero for the other G4S. This case concerns mainly residues H107, L108, I111, R112, K133, F134, K136, and I141. However, caution should be taken to avoid overinterpretation of this result, since 2G10 is not showing any specificity for *h-Telo*.[34] The contact frequency for 2E4 (Figure 4B) shows the emergence of even more defined trends presenting stronger and more localized maxima. In particular, we can mention residues K5, E9, and R12 as well as the regions spanning residues R34, W35, and M46, and residues H67, W68 and R70. However, only a relatively small difference in the interaction patterns between the three G4 s can be highlighted. In particular in the case of *c-Myc* residues Y45, R70, L75, S78, R79, and G80 develop persistent contacts, and hence could be regarded as potential hot-spots for the selectivity of 2E4 towards this G4.

Although, the 2G10/G4 complex involves a larger number of residues developing more persistent contacts than 2EG/G4, this should not be necessarily correlated to a higher affinity towards G4. Indeed, few residues developing stable and persistent interactions may be regarded as more favorable than an extended weakly interacting region. Furthermore, 2G10 is larger than 2E4, thus the higher number of contacting residues may be also regarded as an obvious statistical effect.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

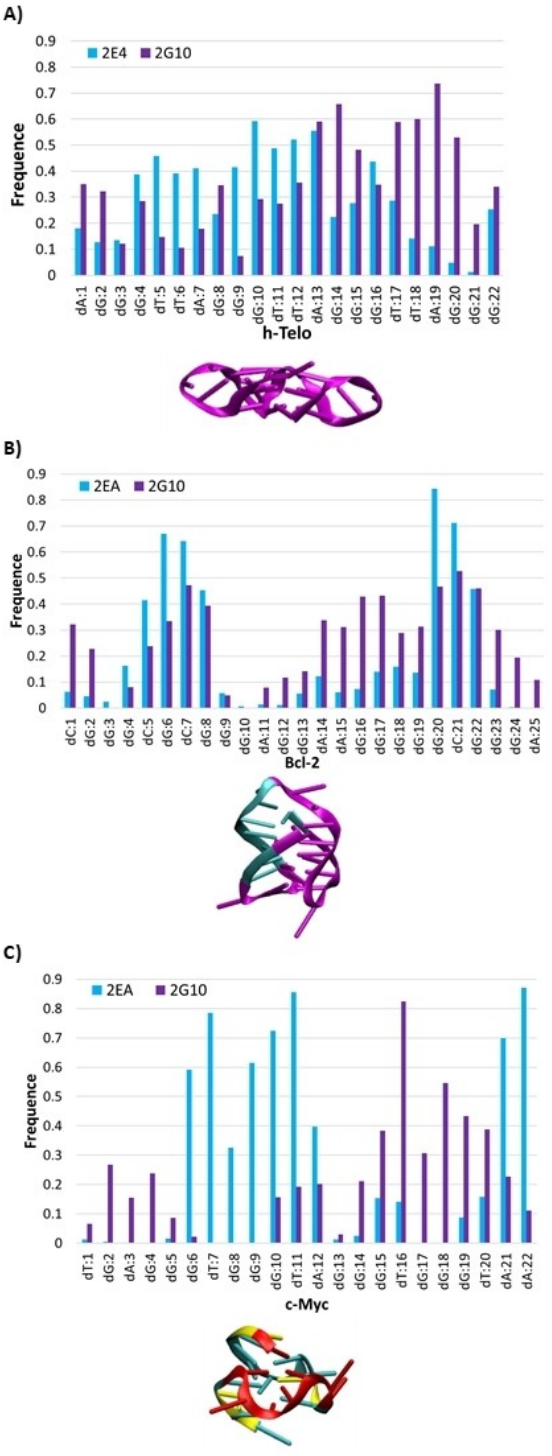


Figure 3. Frequency and location, in magenta, of nucleotide of A) *h-Telo*, B) *Bcl-2* and C) *c-Myc* involved in a 3 Å cutoff from the protein. Note that the color code has been changed for *c-Myc*, because the mode of interaction between the two proteins involves different nucleotides. Thus, the nucleotides most frequently involved with 2E4 are in red, while those interacting preferentially with 2G10 are in yellow.

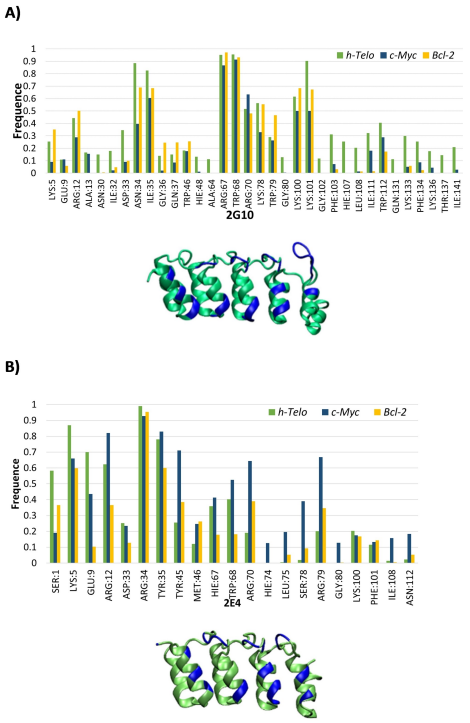


Figure 4. Frequency and location, in blue, of amino acids of A) 2G10 and of B) 2E4 involved in a 3 Å cutoff of G-quadruplex.

Alignment between 2E4 and a c-Myc-specific peptide reveals no sequence similarity

Several examples of peptides able to selectively bind G4 s are reported in the literature. Usually, they are derived from the DHX36 helicase whose α-helix provides the binding interface with the nucleic acid.[38,39] Notably, Minard et al.[40] designed a specific peptide, DM102 (*PGHLKGRRIGLWYASKQGQKNK*), which is able to preferentially recognize G4 in the *c-Myc* promoter. Since 2E4 is also specific to *c-Myc*, it is legitimate to ask whether there is a sequence similarity between DM102 and 2E4. It is also important to note that the artificial peptide DM102 has a hydrocarbon staple (i, $i+7$) between residues R8 and S15, which enforces a defined α-helix. Hence, in addition to sequence similarity, one must also look for stable α-helical secondary structure, which is, indeed, a structural motif frequently present in DARPins. Two algorithms were used: Clustal Omega and M-coffee (see Figure S37). Here it is important to specify that Clustal Omega is an individual method of alignment, while M-coffee is an algorithm that combines results from several individual methods.[41]

Interestingly, the two algorithms show divergent results. Clustal Omega previews similarity mainly concerning the 2E4 region spanning M46 to V62. The representation of the 3D structure of this region (see Figure S38) reveals a helical arrangement, which could partially support the 2E4 selectivity conditions. However, this region is also common to all DARPins,

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

except for residue 58 (residue 70 following Scholz et al. notation),[34] which is embedded in the similarity region, and residues 45 and 46, which border it. Furthermore, the mutated residues at position 46 and 58 are structurally very distant, suggesting a low quality of the alignment. Conversely, M-Coffee alignment highlights three subunits. Two of them have no significance, the first being located at the previously invalidated region, and the third pertaining to the N-terminal region common to all DARPins. Instead, the second subunit is aligned with the R70-R79 region of DM102, as visually represented in Figure 5 by the transparent shaded area. This observation is also coherent with our MD simulations which indicate an increase of the DARPin/*c-Myc* contact frequency for the residues belonging to this region. Furthermore, from a structural point of view, the R70-R79 region is organized in α-helix motif and is located towards the canonical recognition zone of DARPins. Yet, this region is highly conserved among the DARPins designed by Scholz et al.[34] and only the residues bordering the helix, that is, R70, S78, and R79, have been mutated. Indeed, when M-Coffee alignment between DM102 and 2G10 the same DARPin region is evidenced (see Figure S39).

Thus, the search for a conserved sequence between DM102 and 2E4, does not unambiguously justify the selectivity of the DARPin. Going a step further, this could suggest that the recognition of *c-Myc*'s by 2E4 does not necessarily involve sequence similarity between DM102 and 2E4. This is also supported by the fact that the mutations of the wild type sequence as performed by Scholz et al.[34] are mainly concentrated on the peripheral protein loops. Consequently, we decided to focus on structural features which should add up to the rather modest sequence effects and, ultimately, drive the selectivity.

2E4 recognizes a particular structural motif of c-Myc

From our MD simulations two most important factors should be considered when analyzing the local structural arrangements of the DARPin/G4 the contact region. First, the DARPin canonical interaction zone is not consistently interacting throughout the whole MD simulation. Instead, as highlighted in Figure 4(B), other amino acids either located in α-helixes or in peripheral loops develop more persistent interactions.

Furthermore, the analysis of the interaction networks shows that a DARPins/G4 complex is mainly stabilized by electrostatic interaction between positively charged amino acids and the negatively charged backbone of the nucleic acid. The paper by Scholz et al.[35] clearly excludes any interaction between DARPins and canonical double strand DNA. This fact also confirms that the recognition of the nucleic acid should involve important and specific structural motifs, as confirmed by our study. In addition, π-cation interactions are also present mainly when the extended conjugated system of a tetrad faces the DARPin. Even if this interaction appears persistent along the MD simulation it should be confirmed by using quantum chemistry-based modeling, or even hybrid quantum/classical approaches, to avoid any spurious force field artifact and precisely calculate energy interaction terms. However, such a study, even if highly interesting would be out of the scope of the present contribution. Finally, DARPin associates with G4 through its canonical interaction zone involving the α-helix, but also via interactions mediated by the peripheral loops. The interaction with the loop is most pronounced, but not unique, in the case of *h-Telo*, which in the course of the MD simulation departs from the initial docking pose and slides over the DARPin surface until an interaction between its quartet and the peripheral loops is established at around 150 ns (Figure 6). Interestingly, the electrostatic interactions involving the G4 backbone take place mainly through the G4 external loops rather than the tetrad core. However, this conformation appears as scarcely stable, and as a matter of fact the G4 oscillates and reverts to a more classic interaction mode involving one of its accessible quartets. These observations are also consistent with the frequency distribution reported in Figure 3 and explain the specific behavior of *h-Telo*, which due to its high mobility spans different interaction poses and develops rather non-specific contacts with a high number of 2E4 and 2G10 residues.

Thus, the interaction mode involving a quartet is not leading to a specific recognition mode. Hence, interactions between *c-Myc* or *Bcl-2* and DARPin which would be driven by the G4 s quartets (Figure 7) will most probably be trapped in a non-specific recognition and cannot be used to infer on the specific recognition. On the contrary, specificity may be established when DARPins interact mainly with the nucleic acid backbone. The behavior of *Bcl-2*'s, which interacts in a similar non-specific way with 2E4, confirms nicely this statement. Indeed, despite different initial conditions, the G4 again positions itself exposing a quartet to the 2E4 DARPin interaction region. On the contrary, the interaction with 2G10 leads to the exposure of the nucleic acid backbone to the contact region of DARPin and hence, to a selective recognition. As a matter of fact, these results are also coherent with the contact frequency analysis showing that *Bcl-2* interaction with 2G10 is mostly driven by highly conserved and persistent amino acids.

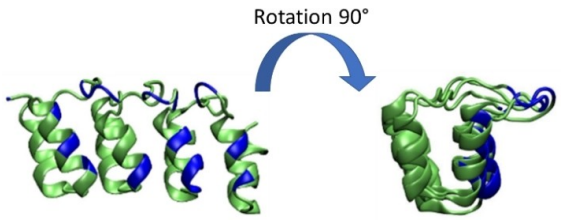


Figure 5. The preferential interaction region of the 2E4 protein is located at the α-helix and the loops and evidenced in blue.

2E4 recognizes a peculiar structural motif of c-Myc

c-Myc is the G4 more consistently promoting an interaction via its backbone (Figure 8). This, in turn, could also point to a greater specificity of its recognition, although *c-Myc*'s is able to interact via its backbone with both 2E4 and 2G10. Thus, to

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

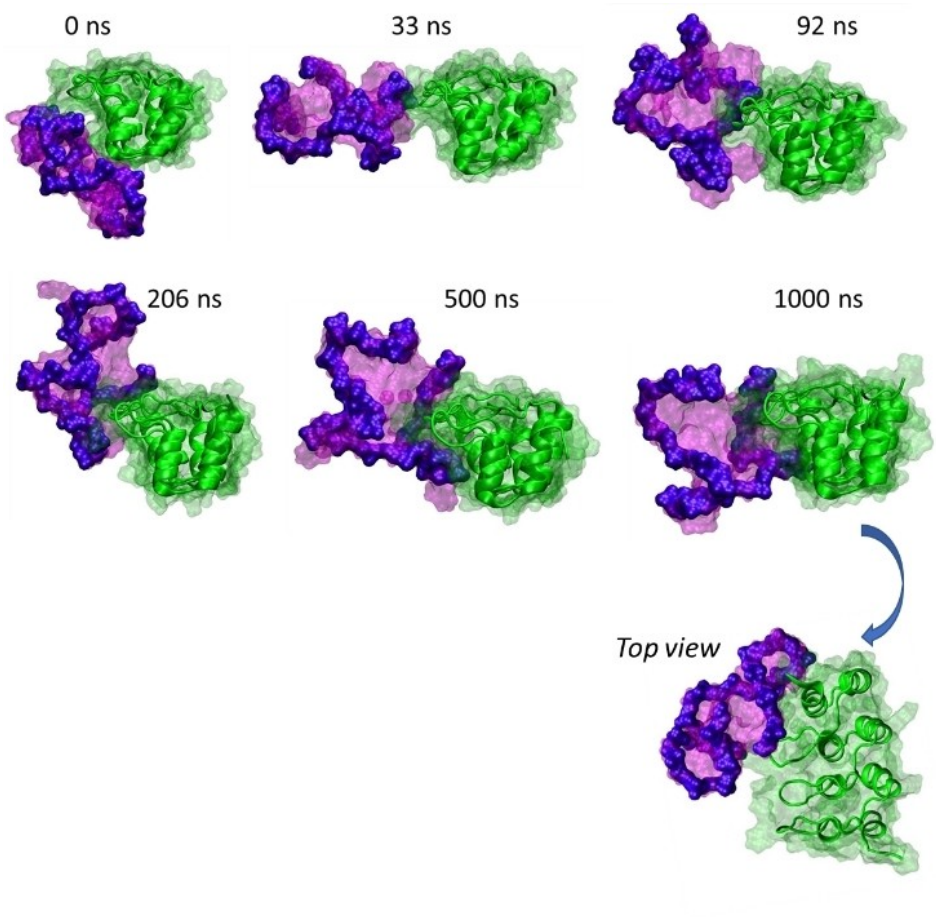


Figure 6. Screenshots of the dynamics of a 2E4/*h-Telo* complex. The *h-Telo* quartet is initially oriented towards the a-helix. But within the first 150 ns, the G-quadruplex rotates and then reorients to interact with the protein loops. During the dynamics, the movements of the G-quadruplex do not affect the protein/G-quartet loops interaction.

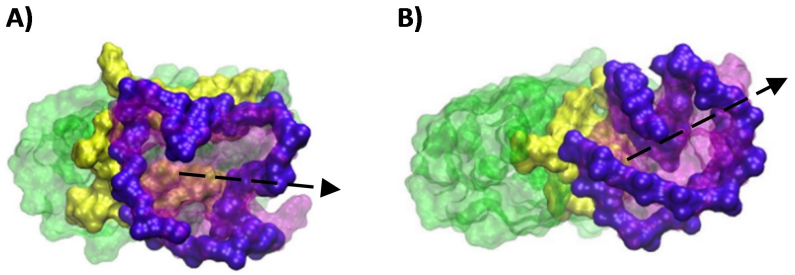


Figure 7. Non-specific 'nucleotide face' interaction found in cluster of the A) 2E4/*h-Telo* pose 1–1 (77.23 % of the MDs) and B) 2E4/*Bcl-2* pose 1–1 (79.01 % of the MDs) complexes.

further justify the selectivity of 2E4 a structural motif specifically recognized by this DARPin should be identified.

The main factor that could lead to a recognized structural motif includes the presence of a backbone folding involving the nucleotides most frequently in contact with the DARPin. This feature can be easily assessed by clustering the MD simulation while checking the maintenance of the interaction patterns in the most populated clusters. By highlighting the E24 highest frequency contact nucleotides, i.e., G6 to A12, A21, and A22, we see that they are involved in the interaction with the protein for

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

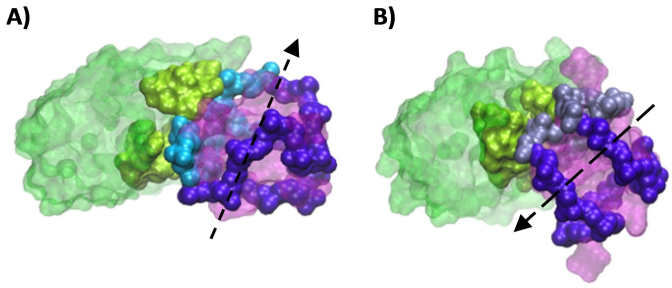


Figure 8. *c-Myc* G-quadruplex interacts preferentially through its backbone A) in the dynamics resulting from the most stable docking pose and B) in the dynamics resulting from a less favorable pose in which the G-quadruplex reorients itself to interact in a pose like the most stable pose found by docking. The previously identified high frequency nucleotides are colored A) in cyan and B) in steel blue, respectively.

the two most important clusters (Figures 8 and 9A, B). However, the two clusters differ by a rotation of about 180° of the G4 on the protein surface (pose 1–1: 78.47 % of the MDs and pose 6–4: 79.55 % of the MDs), as shown in Figure 10. Nonetheless, a well conversed structural motif is evidenced, determined by the folding of the G4 backbone into a U-shaped loop with an extruded nucleotide, further completed by a horizontal extension to the right, and overlaid by a dangling segment (Figure 9). Interestingly, all the structural characteristics are well evidenced in the most populated cluster, while in the secondary structures their identification remains more elusive. Indeed, if the linear extension remains evident, as well as the extruded nucleotide, the U-shaped loop and the appendix are more scarcely visible.

The amino acids located in the interaction site are also conserved between the two most populated clusters (Figure 10). Residues Y45, R70, and R79 organizes around the loop at an average distance of 3 Å, while Y35 is oriented towards the appendix and R34 points towards the linear extension. At a slightly higher distance of around 5 Å, M46 is interacting with the U-shaped loop, W68 is oriented towards the appendix, while R12 and D33 flank the linear extension. In addition, S78 stays close to the extruded nucleotide, probably assuring a further stabilization.

By superposing the two most populated clusters of the 2E4/*c-Myc* complexes a similar positioning of the G4 on the protein site is also observed, which is again consistent with the recognized backbone-based structural motif (Figure 10D). Finally, the increase of the contact frequency observed in the analysis of the MD simulations correlates well with the specific recognition of *c-Myc* by 2E4. Indeed, the region presenting the highest increase in the contact frequency corresponds to the residues recognizing the U-shaped loop and the extruded nucleotide. This, together with the similarity of the interaction pattern found in the superposition of the two G4 s poses, further validates the hypothesis that the selective recognition of *c-Myc* by 2E4 is driven by the structural motif we have identified. Because of this structural-based recognition, and the often-invoked analogy between DARPin and antibodies, it is tempting to characterize this interaction pattern as an epitope/paratope recognition. Here the paratope-like element being the 2E4 interaction site, and the epitope-like region the G4 structural motif identified for *c-Myc*.

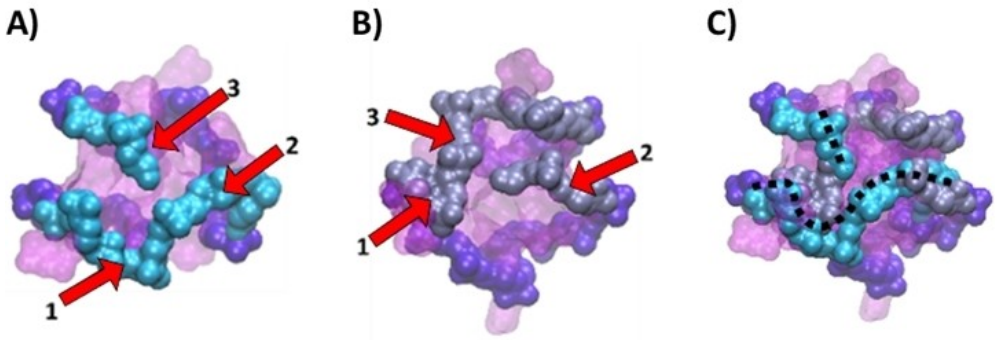


Figure 9. Face view of the *c-Myc* G-quadruplex interacting with 2E4 A) in the dynamics resulting from the most stable docking pose (pose 1–1) and B) in the dynamics resulting from a less favorable pose (pose 6–4): a rotation of 180 degrees with respect to A). Three elements can be identified on each: 1) a U-shaped-loop with an extruded nucleotide, 2) a linear extension and 3) an appendage. The superposition of the two interaction poses C), identify the recognized motif in dotted line.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

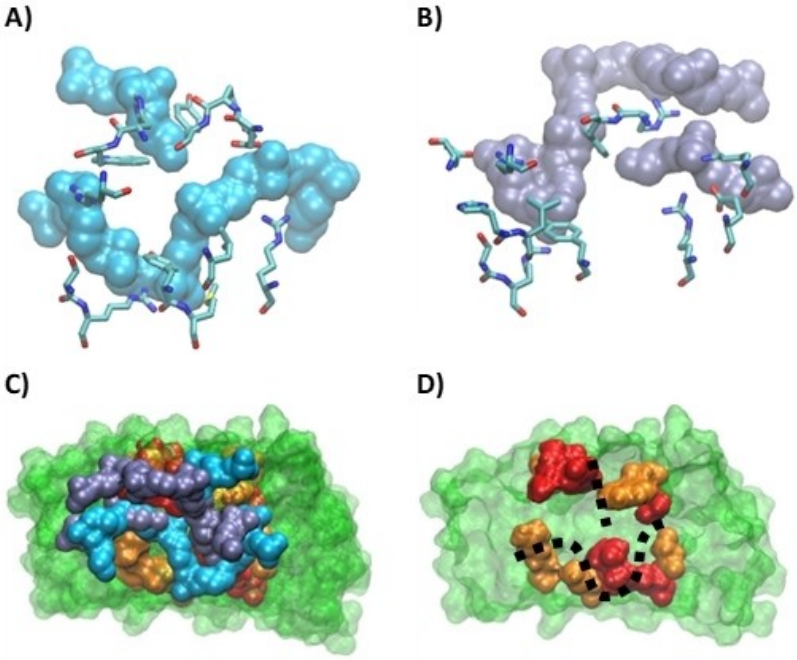


Figure 10. Protein residue involved in the interaction with the c-Myc structural motif A) in the dynamics resulting from the most stable docking pose (pose 1–1) and B) in the dynamics resulting from a less favorable pose (pose 6–4). C) Superposition of the complexes of the two clusters. D) Interaction site of 2E4 with c-Myc; residues at 3 Å are in red and those at 5 Å are in orange.

Conclusion

Our results highlight two modes of interaction for DARPins/G4 complexes. The first one is a non-specific recognition that is established when G4 interacts through its guanine tetrad, or through peripheral nucleotides π-stacked with the tetrads. The second binding mode is driven by the specific recognition of the conformation of the G4 backbone and leads to a DARPin/G4 paratope/epitope like recognition. This specific mode, which we have identified for the 2E4/*c-Myc* complex is based on a peculiar folding motif of the G4 backbone and presents a U-shaped loop with a linear extension and an overhanging short appendix. Consequently, a large extension of the U-shaped loop, also including extruded nucleotides should enhance selective recognition of the G4 s. Conversely, the identification of backbone-based recognition motifs could also improve the rational design of DARPins. Indeed, the quest for selective G4 ligands has a tremendous significance, especially in the proposition of specific anticancer or antiviral agents. Our results, and the first identification of paratope/epitope specific structural recognition may lead to significant development in the design of potentially therapeutics peptides targeting specific G4 arrangements.

Experimental Section

Structure of G4 and reconstruction of DARPins: The structure of the *h-Telo* G4 was retrieved from PDB data bank 1KF1,[42] as well as that of the *c-Myc* (1XAV)[43] and *Bcl-2 (*6ZX7).[44] The sequences of 2E4 and 2G10 DARPins were obtained from the Supporting Information of Ref. [34] and their structure reconstructed with the SWISS-MODEL server.[45] 2E4 were reconstructed based on high similarity with the PDB entry 2CH4[46] and 2G10 was reconstructed on the basis of the 1SVX structure[47] similarity.

Sequence alignment. DM102 peptide and DARPins sequences were aligned using the Clustal Omega on EBI server[48] and the M-coffee server,[49] using their default parameters.

Docking and selection of initial structures: The reconstructed DARPins and G4 were loaded onto the HADDOCK server[50] to perform protein/nucleic acid docking while searching the entire protein and the entire G4 structure and using the standard HADDOCK parameters. Three poses were selected from the docking results (Figure 11), always including the most favorable one. The selection was based on the relative position of G4 with respect to the DARPin. The three poses correspond to an interaction with the G4 backbone, an interaction with the tetrads facing the nucleotides and an interaction developed in a peripheral region of the DARPin. This choice allowed to assure a significant sampling of a complex conformational space, also including rather unfavorable interaction areas, such as the one corresponding to the peripheral binding.

Molecular dynamics simulations: MD simulations has been performed for 2E4 and 2G10 interacting with *c-Myc*, *Bcl-2* and *h-Telo*. Three poses for each complex have been used as starting conditions. Each system was calculated in two independent replicates of 1 μs each, thus a total of 36 simulations of DARPin/G4

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

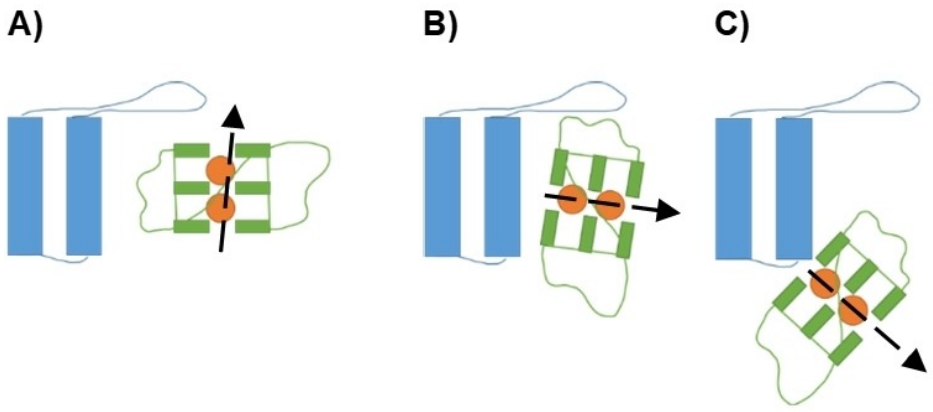


Figure 11. Schematic representation of the three types of poses selected from the docking. There is interaction type A) backbone, B) nucleotide face and C) protein bottom zone.

complexes have been performed. In addition. simulations of the free G4 s and DARPin have also been obtained as a control. All simulations have been run using the NAMD software[51,52] with the Amber parm99 force field[53] including the bsc1 corrections[54] for nucleic acids. A truncated octahedral box of TIP3P[55] water was used to solvate the systems, using periodic boundary condition (PBC). All the calculations were performed in the isothermal and isobaric (NPT) ensemble at a temperature of 300 K and a pressure of 1 atm. A minimal concentration of K^+ ions was added to assure charge equilibration. Hydrogen Mass Repartitioning (HMR)[56] was consistently used, allowing, in combination with the Rattle and Shake algorithms,[57] a timestep of 4 fs to integrate Newton's equations of motion. Finally, the trajectories were analyzed and visualized with VMD,[58] as well as a dedicated script to retrieve G4 structural parameters,[59] while CPPTRAJ[60] was used for clustering.

Acknowledgements

All the calculations were performed on the local cluster of the LPCT and on the regional cluster ExpLor of the University of Lorraine. CNRS and Université de Lorraine and Université Paris Cité, France, and Università di Palermo and from the Ministero dell'Università e Ricerca Scientifica e Tecnologica, are thanked for their support. T.M. thanks University of Palermo for funding a joint Ph.D. program. A.M. thanks ANR (Agence Nationale de la Recherche) and CGI (Commissariat à l'Investissement d'Avenir) for their financial support of this work through Labex SEAM (Science and Engineering for Advanced Materials and devices) ANR 11 LABX 086 and ANR 11 IDEX 05 02. The support of the IdEx "Université Paris 2019" ANR-18-IDEX-0001 and of the Platform P3MB is gratefully acknowledged.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: *c-Myc* promoter · DARPin · epitope/paratope recognition · guanine quadruplex · molecular dynamics

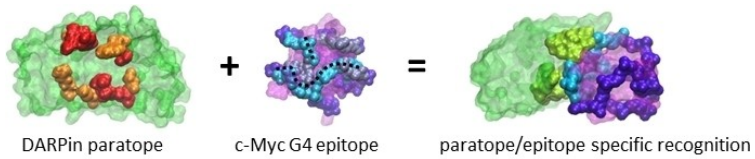
[1] H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
[2] M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770–780.
[3] C. K. Kwok, C. J. Merrick, *Trends Biotechnol.* **2017**, *35*, 997–1013.
[4] J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.* **2020**, *2*, 123–136.
[5] M. García-Díaz, D. Sánchez-García, J. Soriano, M. L. Sagristà, M. Mora, Á. Villanueva, J. C. Stockert, M. Cañete, S. Nonell, *MedChemComm* **2011**, *2*, 616–619.
[6] R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, et al., *Nat. Genet.* **2016**, *48*, 1267–1272.
[7] J. F. Moruno-Manchon, P. Lejault, Y. Wang, B. McCauley, P. Honarpisheh, D. A. M. Scheihing, S. Singh, W. Dang, N. Kim, A. Urayama, et al., *eLife* **2020**, *9*, 10.7554/eLife.52283.
[8] R. Simone, P. Fratta, S. Neidle, G. N. Parkinson, A. M. Isaacs, *FEBS Lett.* **2015**, *589*, 1653–1668.
[9] G. Biffi, D. Tannahill, J. Miller, W. J. Howat, S. Balasubramanian, *PLoS One* **2014**, *9*, 10.1371/journal.pone.0102711.
[10] J. Carvalho, J. L. Mergny, G. F. Salgado, J. A. Queiroz, C. Cruz, *Trends Mol. Med.* **2020**, *26*, 848–861.
[11] Y.-Z. Xu, P. Jenjaroenpun, T. Wongsurawat, S. D. Byrum, V. Shponka, D. Tannahill, E. A. Chavez, S. S. Hung, C. Steidl, S. Balasubramanian, et al., *NAR Cancer* **2020**, *2*, 10.1093/narcan/zcaa029.
[12] N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, *Mol. Cancer* **2021**, *20*, 10.1186/s12943-021-01328–4.
[13] N. Kumari, S. V. Vartak, S. Dahal, S. Kumari, S. S. Desai, V. Gopalakrishnan, B. Choudhary, S. C. Raghavan, *iScience* **2019**, *21*, 288–307.
[14] S. Bielskutel, J. Plavec, P. Podbevšek, *J. Am. Chem. Soc.* **2019**, *141*, 2594–2603.
[15] A. M. Fleming, J. Zhu, Y. Ding, C. J. Burrows, *ACS Chem. Biol.* **2017**, *12*, 2417–2426.
[16] C. C. Hognon, T. Miclot, C. G. Iriepa, A. Francés-Monerris, S. Grandemange, A. Terenzi, M. Marazzi, G. Barone, A. Monari, C. Garcia-Iriepa, et al., *J. Phys. Chem. Lett.* **2020**, *11*, 5661–5667.
[17] T. Miclot, C. Hognon, E. Bignon, A. Terenzi, M. Marazzi, G. Barone, A. Monari, *J. Phys. Chem. Lett.* **2021**, *12*, 10277–10283.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202201824
Chemistry Europe
European Chemical Societies Publishing

[18] A. Francés-Monerris, C. Hognon, T. Miclot, C. García-Iriepa, I. Iriepa, A. Terenzi, S. Grandemange, G. Barone, M. Marazzi, A. Monari, *J. Proteome Res.* **2020**, *19*, 4291–4315.
[19] E. Ruggiero, S. N. Richter, *Annu. Rep. Med. Chem.* **2020**, *54*, 101–131.
[20] C. Zhao, G. Qin, J. Niu, Z. Wang, C. Wang, J. Ren, X. Qu, *Angew. Chem. Int. Ed.* **2021**, *60*, 432–438; *Angew. Chem.* **2021**, *133*, 436–442.
[21] S. R. Wang, Q. Y. Zhang, J. Q. Wang, X. Y. Ge, Y. Y. Song, Y. F. Wang, X. D. Li, B. S. Fu, G. H. Xu, B. Shu, et al., *Cell Chem. Biol.* **2016**, *23*, 1113–1122.
[22] N. Saranathan, P. Vivekanandan, *Trends Microbiol.* **2019**, *27*, 148–163.
[23] M. C. Chen, R. Tippana, N. A. Demeshkina, P. Murat, S. Balasubramanian, S. Myong, A. R. Ferré-D'amaré, *Nature* **2018**, *558*, 465–483.
[24] Z. Pietras, M. A. Wojcik, L. S. Borowski, M. Szewczyk, T. M. Kulinski, D. Cysewski, P. P. Stepien, A. Dziembowski, R. J. Szczesny, *Nat. Commun.* **2018**, *9*, 10.1038/s41467-018-05007–9.
[25] J. C. Darnell, K. B. Jensen, P. Jin, V. Brown, S. T. Warren, R. B. Darnell, *Cell* **2001**, *107*, 489–499.
[26] G. Farine, C. Migliore, A. Terenzi, F. Lo Celso, A. Santoro, G. Bruno, R. Bonsignore, G. Barone, *Eur. J. Inorg. Chem.* **2021**, 1332–1336.
[27] V. Dhamodharan, P. I. Pradeepkumar, *ACS Chem. Biol.* **2019**, *14*, 2102–2114.
[28] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, et al., *Cancer Res.* **2009**, *69*, 7653–7661.
[29] D. Dutta, M. Debnath, D. Müller, R. Paul, T. Das, I. Bessi, H. Schwalbe, J. Dash, *Nucleic Acids Res.* **2018**, *46*, 5355–5365.
[30] S. M. Javadekar, N. M. Nilavar, A. Paranjape, K. Das, S. C. Raghavan, *DNA Res.* **2020**, *27*, 10.1093/dnares/dsaa024.
[31] A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. Brad Johnson, R. M. Brosh, D. Sen, P. M. Lansdorp, *Nucleic Acids Res.* **2017**, *45*, 6252.
[32] H. Y. Liu, Q. Zhao, T. P. Zhang, Y. Wu, Y. X. Xiong, S. K. Wang, Y. L. Ge, J. H. He, P. Lv, T. M. Ou, et al., *Cell Chem. Biol.* **2016**, *23*, 1261–1270.
[33] H. Fernando, R. Rodriguez, S. Balasubramanian, *Biochemistry* **2008**, *47*, 9365–9371.
[34] O. Scholz, S. Hansen, A. Plückthun, *Nucleic Acids Res.* **2014**, *42*, 9182–9194.
[35] P. R. Mittl, P. Ernst, A. Plückthun, *Curr. Opin. Struct. Biol.* **2020**, *60*, 93–100.
[36] A. Plückthun, *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 489–511.
[37] M. T. Stumpp, H. K. Binz, P. Amstutz, *Drug Discovery Today* **2008**, *13*, 695–701.
[38] B. Heddi, V. V. Cheong, H. Martadinata, A. T. Phan, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9608–9613.
[39] M. Y. Yaneva, V. V. Cheong, J. K. Cheng, K. W. Lim, A. T. Phan, *Biochem. Biophys. Res. Commun.* **2020**, *531*, 62–66.
[40] A. Minard, D. Morgan, F. Raguseo, A. Di Porzio, D. Liano, A. G. Jamieson, M. Di Antonio, *Chem. Commun.* **2020**, *56*, 8940–8943.
[41] I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, *Nucleic Acids Res.* **2006**, *34*, 1692–1699.
[42] G. Parkinson, M. Lee, S. Neidle, *Exp. Mol. Pathol.* **2002**, *417*, 876–880.
[43] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
[44] S. Bielskute, J. Plavec, P. Podbevšek, *Nucleic Acids Res.* **2021**, *49*, 2346–2356.
[45] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, et al., *Nucleic Acids Res.* **2018**, *46*, W296-W303.
[46] O. Grubisha, M. Kaminska, S. Duquerroy, E. Fontan, F. Cordier, A. Haouz, B. Raynal, J. Chiaravalli, M. Delepierre, A. Israël, et al., *J. Mol. Biol.* **2010**, *395*, 89–104.
[47] H. K. Binz, P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grütter, A. Plückthun, *Nat. Biotechnol.* **2004**, *22*, 575–582.
[48] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, et al., *Nucleic Acids Res.* **2019**, *47*, W636-W641.
[49] S. Moretti, F. Armougom, I. M. Wallace, D. G. Higgins, C. V. Jongeneel, C. Notredame, *Nucleic Acids Res.* **2007**, *35*, 10.1093/nar/gkm333.
[50] G. C. P. Van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. Van Dijk, S. J. De Vries, A. M. J. J. Bonvin, *J. Mol. Biol.* **2016**, *428*, 720–725.
[51] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.
[52] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, et al., *J. Chem. Phys.* **2020**, *153*, 044130.
[53] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[54] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, et al., *Nat. Methods* **2016**, *13*, 55–58.
[55] P. Mark, L. Nilsson, *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
[56] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.
[57] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952–962.
[58] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
[59] V. Tsvetkov, G. Pozmogova, A. Varizhuk, *J. Biomol. Struct. Dyn.* **2016**, *34*, 705–715.
[60] V. V. Shitov, N. A. Semenov, N. Y. Gozman, *Telecommun. Radio Eng. (English Transl. Elektrosvyaz Radiotekhnika)* **1984**, *38–39*, 14–16.

Manuscript received: June 13, 2022
Accepted manuscript online: July 6, 2022
Version of record online: ■■, ■■■

RESEARCH ARTICLE



DARPins are proteins able to specifically recognize G-quadruplexes. Their small size combined with their easy design makes them a good competitor to antibodies for the identification and localization of G-quadruplex.

Using molecular dynamics calculations, we show that the selectivity of DARPins towards G-quadruplexes is achieved by the recognition of a structural motif adopted by the DNA filament folding.

T. Miclot, Dr. E. Bignon, Prof. A. Terenzi, Prof. S. Grandemange, Prof. G. Barone*, Prof. A. Monari**

1 – 11

G-Quadruplex Recognition by DARPIns through Epitope/Paratope Analogy

6

Conclusions

It is not enough to strike the ear and occupy the eyes, it is necessary to act on the soul.

Discours sur le style - Georges-Louis Leclerc, comte de Buffon - 1753 – Personal translation

G-quadruplex structure and dynamics represent a rich and varied research topic with potential applications ranging from medicine to biotechnology. These unique nucleic acid structures are the subject of much research and multiple national and international projects have emerged. The aim is not only to understand the biological mechanisms involving G-quadruplexes, but also to use them in innovative technologies. In this perspective, the works presented in this manuscript, realized during three years of thesis, aims at deepening the knowledge of G-quadruplex structural complexity and the related properties. In particular, we focus on different aspects of G-quadruplex, ranging from understanding their structural stability to their interactions with proteins.

The results obtained revealed that the use of multi-approach strategy allows to propose sounding models of G-quadruplex. In details, a sequence homology search allows to find the experimentally resolved structure of a G-quadruplex with similarity with the target. The latter is then used as a basis to reconstruct the G-quadruplex structure of interest. This model is then subjected to molecular dynamics simulations to find its conformational landscape. Finally, QM/MM calculations can be performed to obtain the corresponding theoretical circular dichroism spectrum, which can be easily compared with the experimental one. The structural model of G-quadruplex is valid if there is a good matching between the calculated and the experimental spectrum.

8-oxo-guanine or strand breaks have been introduced in the structure of telomeric G-quadruplex, *h-Telo*. These are two types of damage caused by reactive oxygen species and ionizing radiation, respectively. 8-oxo-guanine and strand breaks were introduced according to the same pattern: singularly or in clusters and at different positions. The trajectories resulting from molecular dynamics highlight the high structural stability of G-quadruplexes, even in poresence of lesions. Indeed, the introduction of strand breaks in the loops or in the quartets does not perturb the structural parameters or the global conformlation of the G-quadruplex. The introduction of one 8-oxo-guanine lesion slightly perturbs the structural parameters, but the overall structure remains stable. Only the

introduction of two 8-oxo-guanines can significantly distort the G-quadruplex structure and, in few cases, disrupt it.

The importance of G-quadruplex RNA in the dimerization of the SUD protein of SARS-CoV-2 was also investigated. The analysis of the molecular dynamics evolution shows two possible binding modes. The first involves a single monomer of the protein and G-quadruplex. The other binding mode involves two dimers of the protein for which the G-quadruplex is positioned on both monomers. A free energy calculation on the SUD/G-quadruplex complex shows that the dimeric interaction is the most stable and favors the assembly of the SUD protein in its active conformation. The understanding of this mechanism is very important and provides clues for the development of a possible treatment to fight effectively against the viral infection.

The specific recognition of the G-quadruplex of the c-Myc promoter by the DARPin 2E4 protein was finally. Previous experimental studies highlighted this specificity. However, no structural or sequence explanation has been proposed. As a consequence, the application of molecular docking followed by molecular dynamics simulations on several possible positions pointed out that the specificity of 2E4 towards *c-Myc* is due to the recognition of a structural motif adopted by this G-quadruplex.

Finally, the work presented in this PhD manuscript is part of a research effort enriching the knowledge on G-quadruplexes. It highlights the high resistance of these particular structures, as well as the importance and specificity of their interactions with proteins. Yet the perspectives of this work can be pushed further. Such as establishing the role of alkali metal cations in the stabilization of damaged G-quadruplexes, by calculating their interaction energy. Or, sought to perform similar work on the effect of 8-oxo-guanine and strand breaks can be implemented on G-quadruplex RNA. Thus, it will be possible to verify if there is a difference in stability between DNA and RNA. But it is also possible to consider the rational design of DARPins specific to certain G-quadruplex. This will allow to have an efficient material for the rapid identification or the cellular localization of G-quadruplexes. There is no lack of research avenues concerning this type of non-canonical structure of nucleic acids.

Afterwards, I would like to end this work with this few words from the music *Maintenant je sais* by Jean Gabin :

> [...]
> All my youth, I wanted to say « I know ».
> But, the more I searched, the less I knew.
>
> [...]
> I'm still at my window, I look out, and I wonder ?
>
> Now I know, I know that we never know !
>
> [...]
> That's all I know ! But this I know ... !

Personal translation.

7

Appendices

APPENDIX

Price – Premio Ricerc@STEBICEF

This Call for Papers is aimed at awarding no. 5 prizes of 500.00 euro each, in accordance with the provisions of the current Ricerc@STEBICEF Prize Regulations. Recipients The Call for Proposals is addressed to young PhD students, post-doctoral fellows, grant holders and contract holders who have distinguished themselves for their scientific merits by publishing, in the year 2021 (01.01.2021 - 31.12.2021), as first name or corresponding STEBICEF author, in journals classified in the first or second quartile, according to the Scopus and WOS indices, uploaded and validated on IRIS.

Awards received following the submission of this article:
Miclot T, Hognon C, Bignon E, Terenzi A, Marazzi M, Barone G & Monari A (2021) *Structure and Dynamics of RNA Guanine Quadruplexes in SARS-CoV-2 Genome. Original Strategies against Emerging Viruses.* J Phys Chem Lett 12: 1027710283

Università degli Studi di Palermo

Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche

Direttore

DIPARTIMENTO DI SCIENZE E TECNOLOGIE BIOLOGICHE CHIMICHE E FARMACEUTICHE (STEBICEF)

Nomina Vincitori Premio Ricerc@ STEBICEF 2022

***

IL DIRETTORE

VISTO lo Statuto dell'Università degli Studi di Palermo;

VISTO il D.R. rep. n. 3956/2021, prot.n. 97036 del 06.10.2021, con il quale il Prof. Vincenzo Arizza è stato nominato Direttore del Dipartimento STeBiCeF, per gli anni accademici 2021/2022, 2022/2023, 2023/2024;

VISTO il Regolamento del Premio Ricerc@STEBICEF, emanato con D.R., rep. n.2602 del 18.06.2021, prot. 63742;

VISTO il Bando Premio Ricerc@STEBICEF 2022, Rep. 103 del 13.04.2022, prot. 2957, per l'assegnazione di n.5 premi da 500,00 euro ciascuno, emanato secondo le prescrizioni del succitato vigente Regolamento;

VISTO in particolare l'Art. 3 del Bando, ai sensi del quale, *"un premio verrà assegnato per una pubblicazione interdisciplinare che coinvolga autori del Dipartimento STEBICEF afferenti a due Aree CUN differenti"*;

VISTO il verbale, assunto al prot.n 4797 del 01.06.2022, relativo alla seduta della Commissione AQ-Ricerca, che ha valutato le istanze di partecipazione pervenute e stilato una graduatoria di merito, ai sensi dell'Art. 3 del suddetto Bando;

DECRETA

di approvare la seguente graduatoria di merito:

1. Dott. Salvatore Emanuele Drago;
2. Dott.ssa Daniela Carbone;
3. Dott.ssa Teresa Faddetta;
4. Dott. Tom Miclot;
5. Dott. Gabriele La Monica;
6. Dott. Morena Anthony.

Risultano pertanto vincitori il Dott. Salvatore Emanuele Drago, la Dott.ssa Daniela Carbone, la Dott.ssa Teresa Faddetta, il Dott. Tom Miclot.

Il Direttore

Prof. Vincenzo Arizza

Viale delle Scienze, Ed. 16 - 90128 Palermo - +39 091 238 97111 - +39 091 238 67515
dipartimento.stebicef@unipa.it - dipartimento.stebicef@cert.unipa.it (pec) - https://www.unipa.it/dipartimenti/stebicef

APPENDIX

Interview

We know what we are, but know not what we may be.

Ophelia, in Hamlet, Act 4, Scene 5. William Shakespeare.

At the beginning of my second year of thesis, Dominique Daloz, director of the C2MP doctoral school, gave me the honor to propose me to realize a profile of my academic career and my PhD thesis. Questions was led by Caroline Sobolewski, PhD Communication Officer at the University of Lorraine.

The interview was published in the University of Lorraine's news journal, named *Factuel*. The following page presents a reproduction of this interview (in french), with the web links.

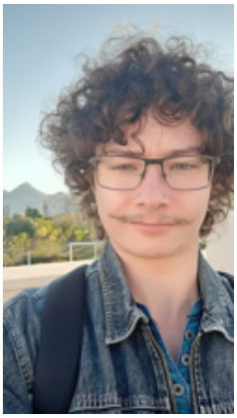
With these few lines, I would like to thank Dominique Daloz and Caroline Sobolewski and convey to them all my consideration.

NOS LABORATOIRES |

[Rencontre] Tom Miclot, doctorant au Laboratoire de physique et chimie théoriques

Publié le 3/11/2020

Tom Miclot, doctorant en deuxième année de thèse, nous présente son parcours et sa thèse réalisée en co-tutelle France-Italie.



Quel est ton parcours ?
Mon parcours d'étude est assez atypique. Après mon Bac technologie en Physique et Chimie de Laboratoire, j'ai choisi de m'orienter vers le BTS Biotechnologies au lycée Stanislas à Villers-Lès-Nancy. C'est lors de ces études que j'ai eu l'opportunité de partir faire un stage de deux mois dans un laboratoire, à l'Université de Limerick (Irlande). Puis, j'ai intégré la Faculté des Sciences de Nancy en 2015 où j'ai obtenu un DEUG et une Licence en Science de la Vie. Ensuite, j'ai continué mon cursus en m'inscrivant au Master Biotechnologies - option Ingénierie Moléculaire. Là, j'ai eu un premier contact avec la recherche théorique grâce à un stage de Master 1 au sein du Laboratoire de Physique et Chimie Théoriques. L'année d'après, j'ai préparé un sujet de mémoire sur les protéines répétées Armadillo, à l'Université de Zurich (Suisse). S'ajoute à cela mon entrée au Master d'Épistémologie, Logique et Histoire des Sciences (MADELHIS) à l'Université des Sciences Humaines et Sociales à Nancy. Je l'ai suivi à distance et parallèlement à la seconde année de mon premier master, ainsi que pendant ma première année de thèse en 2019. Aujourd'hui, je plonge dans ma deuxième année d'une thèse que j'effectue en cotutelle entre le Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche (Università degli Studi di Palermo, Italie), et le Laboratoire de Physique et Chimie Théoriques (FST Nancy, France).

Je suis également étudiant entrepreneur : je suis porteur d'un projet d'entreprise innovante qui est accompagné par le Pôle entrepreneuriat étudiant de Lorraine (PEEL). Nous formons une équipe de cinq personnes et nous travaillons sur une nouvelle technologie visant à éliminer plus facilement les perturbateurs endocriniens présents dans l'eau potable. Bien sûr, ce projet n'est encore qu'à un stade embryonnaire. Vu la portée du sujet, nous espérons le voir se concrétiser d'ici quelque temps.

Sur quelle thématique travailles-tu ?
Mon sujet de thèse porte sur l'étude des lésions des acides nucléiques (ADN et ARN) et de leurs interactions avec des protéines. Afin de mener à bien mes recherches, j'utilise des techniques informatiques qui opèrent comme un « microscope virtuel » permettant d'observer les molécules à l'échelle atomique. Concrètement, j'exploite des structures résolues d'acides nucléiques et de protéines qui proviennent de banques de données telles que la Protein Data Bank ou la Nucleic Acid Database. Ensuite, j'utilise des logiciels pour modifier la structure d'un G-quadruplex (une forme particulière d'ADN à 4 brins) en lui ajoutant des lésions. Puis, j'emploie d'autres outils informatiques capables de donner le comportement des molécules dans de l'eau. En fait, il s'agit de simulations numériques qui modélisent le mouvement des atomes au cours du temps. C'est ce que l'on nomme la dynamique moléculaire. Puis, lorsque cela est nécessaire, je réalise des calculs d'énergie libre sur les modèles issus des dynamiques. Réaliser ce type de calculs permet de connaître les grandeurs d'association entre deux molécules. Par exemple, cela permet de savoir si une protéine s'associe préférentiellement avec un ADN sain, ou avec un ADN lésé.

Pourquoi as-tu décidé de faire une thèse en co-tutelle France-Italie ?
Mes voyages internationaux en Irlande et en Suisse m'ont beaucoup appris, surtout sur la valeur de la Science et le savoir-faire bâti hors de France. Alors, c'est en cosmopolite que j'ai souhaité poursuivre une thèse à l'étranger. D'abord, j'ai pensé à l'Amérique ou à l'Australie. Puis, l'occasion d'une thèse en co-tutelle m'a été présentée par Antonio Monari que je connaissais déjà depuis mon stage de Master 1 sous sa supervision. Je fus donc séduit, car j'y ai vu l'opportunité de prendre part à un projet de recherche s'inscrivant dans une portée européenne et pluridisciplinaire, avec la promesse d'un double diplôme décerné par chaque université.
Outre le haut intérêt scientifique, la thèse en cotutelle est avant tout un échange : elle concrétise l'aspiration justifiée de chercheurs à fonder un projet commun et marque la volonté d'établir et de valoriser une science libre dans une société mondialisée. En fait, la co-tutelle symbolise la détermination de plusieurs acteurs visant à promouvoir l'ouverture de la science, tant à l'échelle locale que dans la dimension internationale.

Que fais-tu en Italie ?
Le LPCT à Nancy est spécialisé en recherche théorique, mais le laboratoire italien dispose d'un équipement expérimental important et me permet d'étudier et de caractériser les structures G-quadruplex. En fait, modèles théoriques et expériences se complètent.

http://factuel.univ-lorraine.fr/node/15443

References

[1] Jocelyn E. Krebs, Elliott S. Goldstein, and Stephen T. Kilpatrick. *Lewin's genes XII*. Burlington, MA: Jones & Bartlett Learning, 2018. ISBN: 9781284104493.

[2] Friedrich Miescher. "Ueber die chemische Zusammensetzung der Eiterzellen". In: *Medicinisch-chemische Untersuchungen: 4*. August Hirschwald, 1871, pp. 441–462. URL: https://books.google.it/books?id=YJRTAAAAcAAJ.

[3] Paul O. P. Ts'o and J. Eisinger. *Basic principles in nucleic acid chemistry*. New York: Academic Press, 1974. ISBN: 9780127019017.

[4] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". In: *Nature* 171.4356 (1953), pp. 737–738. DOI: 10.1038/171737a0.

[5] M. Spencer. "The stereochemistry of deoxyribonucleic acid. II. Hydrogen-bonded pairs of bases". In: *Acta Crystallographica* 12.1 (1959), pp. 66–71. ISSN: 0365-110X. DOI: 10.1107/S0365110X59000160.

[6] K. Hoogsteen. "The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine". In: *Acta Crystallographica* 16.9 (1963), pp. 907–916. ISSN: 0365110X. DOI: 10.1107/S0365110X63002437.

[7] John P. Bartley, Tom Brown, and Andrew N. Lane. "Solution Conformation of an Intramolecular DNA Triplex Containing a Nonnucleotide Linker: Comparison with the DNA Duplex". In: *Biochemistry* 36.47 (1997), pp. 14502–14511. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi970710q.

[8] C. W. Hilbers et al. "The Hairpin Elements of Nucleic Acid Structure: DNA and RNA Folding". In: ed. by Fritz Eckstein and David M. J. Lilley. Vol. 8. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 56–104. ISBN: 9783642786686. DOI: 10.1007/978-3-642-78666-2_4.

[9] A. T. Phan. "Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix". In: *Nucleic Acids Research* 30.21 (2002), pp. 4618–4625. ISSN: 13624962. DOI: 10.1093/nar/gkf597.

[10] Dipankar Sen and Walter Gilbert. "A sodium-potassium switch in the formation of four-stranded G4-DNA". In: *Nature* 344.6265 (1990), pp. 410–414. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/344410a0.

[11] R. Garrett and Charles M. Grisham. *Biochemistry*. 4th ed. Belmont, CA: Brooks/Cole, Cengage Learning, 2010. ISBN: 9780495109358.

[12] P. Yakovchuk. "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix". In: *Nucleic Acids Research* 34.2 (2006), pp. 564–574. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkj454.

[13] James D. Watson, ed. *Molecular biology of the gene*. Seventh edition. Boston: Pearson, 2014. ISBN: 9780321762436.

[14] Takashi Ohyama, ed. *DNA conformation and transcription*. Molecular biology intelligence unit. Georgetown, Tex. : New York, NY: Landes Bioscience ; Springer Science Business Media, 2005. ISBN: 9780387255798.

[15] Romaric Forêt. *Dico de bio*. 3e éd. Bruxelles [Paris]: De Boeck, 2012. ISBN: 9782804171452.

[16] Anthony J. F. Griffiths et al. *Introduction to genetic analysis*. Eleventh edition. New York, NY: W.H. Freeman & Company, a Macmillan Education imprint, 2015. ISBN: 9781464109485.

[17] B Ewald, D Sampath, and W Plunkett. "Nucleoside analogs: molecular mechanisms signaling cell death". In: *Oncogene* 27.50 (2008), pp. 6522–6537. ISSN: 0950-9232, 1476-5594. DOI: 10.1038/onc.2008.316.

[18] Deepa Sampath, V Ashutosh Rao, and William Plunkett. "Mechanisms of apoptosis induction by nucleoside analogs". In: *Oncogene* 22.56 (2003), pp. 9063–9074. ISSN: 0950-9232, 1476-5594. DOI: 10.1038/sj.onc.1207229.

[19] Jay P. Parrish et al. "DNA Alkylation Properties of Yatakemycin". In: *Journal of the American Chemical Society* 125.36 (2003), pp. 10971–10976. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja035984h.

[20] N. Shrivastav, D. Li, and J. M. Essigmann. “Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation”. In: *Carcinogenesis* 31.1 (2010), pp. 59–70. ISSN: 0143-3334, 1460-2180. DOI: 10.1093/carcin/bgp262.

[21] Tod Duncan et al. “Reversal of DNA alkylation damage by two human dioxygenases”. In: *Proceedings of the National Academy of Sciences* 99.26 (2002), pp. 16660–16665. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.262589799.

[22] Eric Krueger et al. “Modeling and Analysis of Intercalant Effects on Circular DNA Conformation”. In: *ACS Nano* 10.9 (2016), pp. 8910–8917. ISSN: 1936-0851, 1936-086X. DOI: 10.1021/acsnano.6b04876.

[23] James E Cleaver. “UV damage DNA repair and skin carcinogenesis”. In: *Frontiers in Bioscience* 7.4 (2002), pp. d1024–1043. ISSN: 10939946, 10934715. DOI: 10.2741/A829.

[24] Thierry Douki and Jean Cadet. “Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions”. In: *Biochemistry* 40.8 (2001), pp. 2495–2501. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi0022543.

[25] Jean-Luc Ravanat, Thierry Douki, and Jean Cadet. “Direct and indirect effects of UV radiation on DNA and its components”. In: *Journal of Photochemistry and Photobiology B: Biology* 63.1-3 (2001), pp. 88–102. ISSN: 10111344. DOI: 10.1016/S1011-1344(01)00206-8.

[26] H Nikjoo et al. “Radiation track, DNA damage and responsea review”. In: *Reports on Progress in Physics* 79.11 (2016), p. 116601. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/79/11/116601.

[27] W D Henner et al. “gamma Ray induced deoxyribonucleic acid strand breaks. 3’ Glycolate termini.” In: *Journal of Biological Chemistry* 258.2 (1983), pp. 711–713. ISSN: 00219258. DOI: 10.1016/S0021-9258(18)33104-1.

[28] Clemens Von Sonntag et al. “Radiation-Induced Strand Breaks in DNA: Chemical and Enzymatic Analysis of End Groups and Mechanistic Aspects”. In: *Advances in Radiation Biology*. Vol. 9. Elsevier, 1981, pp. 109–142. ISBN: 9780120354092. DOI: 10.1016/B978-0-12-035409-2.50009-6.

[29] Tom Miclot et al. “Never Cared for What They Do: High Structural Stability of Guanine-Quadruplexes in the Presence of Strand-Break Damage”. In: *Molecules* 27.10 (2022), p. 3256. ISSN: 1420-3049. DOI: 10.3390/molecules27103256.

[30] Junji Morita et al. “Sequence specific damage of DNA induced by reducing sugars”. In: *Nucleic Acids Research* 13.2 (1985), pp. 449–458. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/13.2.449.

[31] T A Kunkel. “Mutational specificity of depurination.” In: *Proceedings of the National Academy of Sciences* 81.5 (Mar. 1984), pp. 1494–1498. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.81.5.1494.

[32] Bruce K. Duncan and Jeffrey H. Miller. “Mutagenic deamination of cytosine residues in DNA”. In: *Nature* 287.5782 (Oct. 1980), pp. 560–561. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/287560a0.

[33] Steen Steenken and Slobodan V. Jovanovic. “How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution”. In: *Journal of the American Chemical Society* 119.3 (1997), pp. 617–618. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja962255b.

[34] N. R. Jena and P. C. Mishra. “Mechanisms of Formation of 8-Oxoguanine Due To Reactions of One and Two OH Radicals and the H $_2$ O $_2$ Molecule with Guanine: A Quantum Computational Study”. In: *The Journal of Physical Chemistry B* 109.29 (2005), pp. 14205–14218. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp050646j.

[35] Geneviève Pratviel and Bernard Meunier. “Guanine Oxidation: One- and Two-Electron Reactions”. In: *Chemistry - A European Journal* 12.23 (2006), pp. 6018–6030. ISSN: 0947-6539, 1521-3765. DOI: 10.1002/chem.200600539.

[36] William L. Neeley and John M. Essigmann. “Mechanisms of Formation, Genotoxicity, and Mutation of Guanine Oxidation Products”. In: *Chemical Research in Toxicology* 19.4 (2006), pp. 491–505. ISSN: 0893-228X, 1520-5010. DOI: 10.1021/tx0600043.

[37] R. Wood. “Which DNA polymerases are used for DNA-repair in eukaryotes?” In: *Carcinogenesis* 18.4 (1997), pp. 605–610. ISSN: 14602180. DOI: 10.1093/carcin/18.4.605.

[38] Sara K Martin and Richard D Wood. “DNA polymerase ζ in DNA replication and repair”. In: *Nucleic Acids Research* 47.16 (2019), pp. 8348–8361. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz705.

[39] Wen-Jin Wu, Wei Yang, and Ming-Daw Tsai. “How DNA polymerases catalyse replication and repair with contrasting fidelity”. In: *Nature Reviews Chemistry* 1.9 (2017), p. 0068. ISSN: 2397-3358. DOI: 10.1038/s41570-017-0068.

[40] T. Izumi and I. Mellon. “Chapter 17 - Base Excision Repair and Nucleotide Excision Repair”. In: *Genome Stability*. Ed. by Igor Kovalchuk and Olga Kovalchuk. Boston: Academic Press, 2016, pp. 275–302. ISBN: 9780128033098.

[41] Tae-Hee Lee and Tae-Hong Kang. “DNA Oxidation and Excision Repair Pathways”. In: *International Journal of Molecular Sciences* 20.23 (2019), p. 6092. ISSN: 1422-0067. DOI: 10.3390/ijms20236092.

[42] Ralph Scully et al. “DNA double-strand break repair-pathway choice in somatic mammalian cells”. In: *Nature Reviews Molecular Cell Biology* 20.11 (2019). ISSN: 1471-0080. DOI: 10.1038/s41580-019-0152-0.

[43] Michael R. Lieber. “The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway”. In: *Annual Review of Biochemistry* 79 (2010), pp. 181–211. ISSN: 1545-4509. DOI: 10.1146/annurev.biochem.052308.093131.

[44] Nees Jan van Eck and Ludo Waltman. “Software survey: VOSviewer, a computer program for bibliometric mapping”. In: *Scientometrics* 84.2 (2010), pp. 523–538. ISSN: 0138-9130, 1588-2861. DOI: 10.1007/s11192-009-0146-3.

[45] Jessica J. King et al. “DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells”. In: *Journal of the American Chemical Society* 142.49 (2020), pp. 20600–20604. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.0c11708.

[46] Daniela Rhodes and Hans J. Lipps. “G-quadruplexes and their regulatory roles in biology”. In: *Nucleic Acids Research* 43.18 (2015), pp. 8627–8637. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv862.

[47] Paulina Prorok et al. “Involvement of G-quadruplex regions in mammalian replication origin activity”. In: *Nature Communications* 10 (2019), p. 3274. DOI: 10.1038/s41467-019-11104-0.

[48] Katrin Paeschke et al. “Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo”. In: *Nature Structural & Molecular Biology* 12.10 (2005), pp. 847–854. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb982.

[49] Shankar Balasubramanian, Laurence H. Hurley, and Stephen Neidle. “Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?” In: *Nature Reviews Drug Discovery* 10.4 (2011), pp. 261–275. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd3428.

[50] Sefan Asamitsu et al. “Perspectives for Applying G-Quadruplex Structures in Neurobiology and Neuropharmacology”. In: *Int. J. Mol. Sci.* 20.12 (2019), p. 2884. DOI: 10.3390/ijms20122884.

[51] Norifumi Shioda et al. “Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome”. In: *Nature Medicine* 24.6 (2018), pp. 802–813. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-018-0018-6. (Visited on 07/15/2022).

[52] Nandhini Saranathan and Perumal Vivekanandan. “G-Quadruplexes: More than just a kink in microbial genomes”. In: *Trends in microbiology* 27.2 (2019), pp. 148–163.

[53] Mathieu Métifiot et al. “G-quadruplexes in viruses: function and potential therapeutic applications”. In: *Nucleic acids research* 42.20 (2014), pp. 12352–12366.

[54] Emanuela Ruggiero and Sara N Richter. “G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy”. In: *Nucleic acids research* 46.7 (2018), pp. 3270–3283.

[55] Jinzhi Tan et al. “The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes”. In: *PLoS pathogens* 5.5 (2009).

[56] Jinzhi Tan et al. “The SARS-unique domain(SUD) of SARS coronavirus is an oligo (G)-binding protein”. In: *Biochemical and biophysical research communications* 364.4 (2007), pp. 877–882.

[57] Yuri Kusov et al. “A G-quadruplex-binding macrodomain within the SARS-unique domain is essential for the activity of the SARS-coronavirus replication–transcription complex”. In: *Virology* 484 (2015), pp. 313–322.

[58] Chuanqi Zhao et al. “Targeting RNA GQuadruplex in SARSCoV2: A Promising Therapeutic Target for COVID19?” In: *Angewandte Chemie* 133.1 (Jan. 2021), pp. 436–442. ISSN: 0044-8249, 1521-3757. DOI: 10.1002/ange.202011419. (Visited on 07/15/2022).

[59] Shruti Mishra et al. “Guanine Quadruplex DNA Regulates Gamma Radiation Response of Genome Functions in the Radioresistant Bacterium *Deinococcus radiodurans*”. In: *Journal of Bacteriology* 201.17 (Sept. 2019). Ed. by Victor J. DiRita. ISSN: 0021-9193, 1098-5530. DOI: 10.1128/JB.00154-19. (Visited on 07/15/2022).

[60] Vikas Yadav et al. “G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance”. In: *Frontiers in Plant Science* 8 (2017), p. 1163. ISSN: 1664-462X. DOI: 10.3389/fpls.2017.01163.

[61] Xiaofei Yang et al. “RNA G-quadruplex structures exist and function in vivo in plants”. In: *Genome Biology* 21.1 (2020), p. 226. ISSN: 1474-760X. DOI: 10.1186/s13059-020-02142-9.

[62] Emilia Puig Lombardi and Arturo Londoño-Vallejo. “A guide to computational methods for G-quadruplex prediction”. In: *Nucleic Acids Research* 48.1 (2020), pp. 1–15. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1097.

[63] Chun Kit Kwok and Catherine J. Merrick. “G-Quadruplexes: Prediction, Characterization, and Biological Application”. In: *Trends in Biotechnology* 35.10 (2017), pp. 997–1013. ISSN: 01677799. DOI: 10.1016/j.tibtech.2017.06.012. (Visited on 07/15/2022).

[64] Scott L. Forman et al. “Toward Artificial Ion Channels: A Lipophilic G-Quadruplex”. In: *Journal of the American Chemical Society* 122.17 (2000), pp. 4060–4067. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja9925148.

[65] Lucas T. Gray et al. “G-quadruplexes Sequester Free Heme in Living Cells”. In: *Cell Chemical Biology* 26.12 (2019), 1681–1691.e5. ISSN: 24519456. DOI: 10.1016/j.chembiol.2019.10.003.

[66] Nisreen Shumayrikh and Dipankar Sen. “HemeG-Quadruplex DNAzymes: Conditions for Maximizing Their Peroxidase Activity”. In: *G-Quadruplex Nucleic Acids*. Ed. by Danzhou Yang and Clement Lin. Vol. 2035. New York, NY: Springer New

York, 2019, pp. 357–368. ISBN: 9781493996650. DOI: 10.1007/978-1-4939-9666-7_22.

[67] Prabhpreet Singh et al. "G-quartet type self-assembly of guanine functionalized single-walled carbon nanotubes". In: *Nanoscale* 4.6 (2012), p. 1972. ISSN: 2040-3364, 2040-3372. DOI: 10.1039/c2nr11849a.

[68] Loic Stefan and David Monchaud. "Applications of guanine quartets in nanotechnology and chemical biology". In: *Nature Reviews Chemistry* 3.11 (2019), pp. 650–668. ISSN: 2397-3358. DOI: 10.1038/s41570-019-0132-0.

[69] Jean-Louis Mergny and Dipankar Sen. "DNA Quadruple Helices in Nanotechnology". In: *Chemical Reviews* 119.10 (2019), pp. 6290–6325. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.8b00629.

[70] Donald Voet, Judith G Voet, and Guy G Rousseau. *Biochimie*. Ed. by De Boeck. 2010. ISBN: 9782804147952.

[71] N. B. Leontis. "The non-Watson-Crick base pairs and their associated isosteric-ity matrices". In: *Nucleic Acids Research* 30.16 (2002), pp. 3497–3531. ISSN: 13624962. DOI: 10.1093/nar/gkf481.

[72] Jerry Donohue and Kenneth N. Trueblood. "Base pairing in DNA". In: *Journal of Molecular Biology* 2.6 (1960), pp. 363–371. DOI: 10.1016/S0022-2836(60)80047-2.

[73] Evgenia N. Nikolova et al. "Transient Hoogsteen base pairs in canonical duplex DNA". In: *Nature* 470.7335 (2011), pp. 498–502. DOI: 10.1038/nature09775.

[74] Heidi S. Alvey et al. "Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics". In: *Nature Communications* 5.1 (2014), p. 4786. DOI: 10.1038/ncomms5786.

[75] Maxim D. Frank-Kamenetskii and Sergei M. Mirkin. "Triplex DNA Structures". In: *Annual Review of Biochemistry* 64.1 (1995), pp. 65–95. DOI: 10.1146/annurev.bi.64.070195.000433.

[76] Hala Abou Assi et al. "i-Motif DNA: structural features and significance to cell biology". In: *Nucleic Acids Research* 46.16 (2018), pp. 8038–8056. DOI: 10.1093/nar/gky735.

[77] Martin Gellert, Marie N. Lipsett, and David R. Davies. “HELIX FORMATION BY GUANYLIC ACID”. In: *Proceedings of the National Academy of Sciences* 48.12 (1962), pp. 2013–2018. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.48.12.2013.

[78] Stephen Neidle and Shankar Balasubramanian, eds. *Quadruplex nucleic acids*. RSC biomolecular sciences. Cambridge: RSC Pub, 2006. ISBN: 9780854043743.

[79] Jihye Moon et al. “Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin”. In: *Biochemistry and Biophysics Reports* 2 (2015), pp. 29–35. DOI: 10.1016/j.bbrep.2015.03.012.

[80] F. Zaccaria, G. Paragi, and C. Fonseca Guerra. “The role of alkali metal cations in the stabilization of guanine quadruplexes: why K^+ is the best”. In: *Physical Chemistry Chemical Physics* 18.31 (2016), pp. 20895–20904. DOI: 10.1039/C6CP01030J.

[81] Besik I. Kankia and Luis A. Marky. “Folding of the Thrombin Aptamer into a G-Quadruplex with Sr^{2+} : Stability, Heat, and Hydration”. In: *Journal of the American Chemical Society* 123.44 (2001), pp. 10799–10804. DOI: 10.1021/ja010008o.

[82] Eric Largy, Jean-Louis Mergny, and Valérie Gabelica. “Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability”. In: *The Alkali Metal Ions: Their Role for Life*. Ed. by Astrid Sigel, Helmut Sigel, and Roland K. O. Sigel. Vol. 16. Cham: Springer International Publishing, 2016, pp. 203–258. DOI: 10.1007/978-3-319-21756-7_7.

[83] Pascale Hazel et al. “Loop-Length-Dependent Folding of G-Quadruplexes”. In: *Journal of the American Chemical Society* 126.50 (2004), pp. 16405–16415. DOI: 10.1021/ja045154j.

[84] M. Clarke Miller et al. “Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3 Sequence of $d(AG_3(TTAG_3)_3)$”. In: *Journal of the American Chemical Society* 132.48 (2010), pp. 17105–17107. DOI: 10.1021/ja105259m.

[85] Jonathan B. Chaires and David Graves, eds. *Quadruplex nucleic acids*. Topics in current chemistry 330. Springer Verlag, 2013. ISBN: 9783642347429.

[86] Daisuke Miyoshi, Hisae Karimata, and Naoki Sugimoto. “Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions”. In: *Journal of the American Chemical Society* 128.24 (2006), pp. 7957–7963. DOI: 10.1021/ja061267m.

[87] Mateus Webba da Silva et al. “Design of a G-Quadruplex Topology through Glycosidic Bond Angles”. In: *Angewandte Chemie International Edition* 48.48 (2009), pp. 9167–9170. DOI: 10.1002/anie.200902454.

[88] Sarah Burge et al. “Quadruplex DNA: sequence, topology and structure”. In: *Nucleic Acids Research* 34.19 (2006), pp. 5402–5415. DOI: 10.1093/nar/gkl655.

[89] Markus Meier et al. “Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge”. In: *Nucleic Acids Research* 46.10 (2018), pp. 5319–5331. DOI: 10.1093/nar/gky307.

[90] Vineeth Thachappilly Mukundan and Anh Tuân Phan. “Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences”. In: *Journal of the American Chemical Society* 135.13 (2013), pp. 5017–5028. DOI: 10.1021/ja310251r.

[91] Poulomi Das et al. “Bulges in left-handed G-quadruplexes”. In: *Nucleic Acids Research* 49.3 (2021), pp. 1724–1736. DOI: 10.1093/nar/gkaa1259.

[92] Christopher Jacques Lech, Brahim Heddi, and Anh Tuân Phan. “Guanine base stacking in G-quadruplex nucleic acids”. In: *Nucleic Acids Research* 41.3 (2013), pp. 2034–2046. DOI: 10.1093/nar/gks1110.

[93] Colin A. Ronan. *Histoire mondiale des sciences*. 2. éd. Points Sciences 129. Paris: Éd. du Seuil, 2000. ISBN: 978-2-02-036237-5.

[94] Richard Feynman, Matthew Sands, and Robert B. Leighton. *The Feynman Lectures on Physics Vol. II Ch. 1: Electromagnetism*. Caltech. URL: https://www.feynmanlectures.caltech.edu/II_01.html (visited on 02/16/2021).

[95] Viorel Sergiesco. *CHAMP, physique*. URL: https://www.universalis.fr/encyclopedie/champ-physique/ (visited on 02/16/2021).

[96] Jean-Claude Maurizot. “Particules élémentaires et interactions fondamentales”. In: *Techniques de l’ingénieur*, R6470 V2 (2009).

[97] James Clerk Maxwell. “VIII. A dynamical theory of the electromagnetic field”. In: *Royal Society* 155 (1865). DOI: https://doi.org/10.1098/rstl.1865.0008.

[98] Michel Ney. “Bases de lélectromagnétisme”. In: *Techniques de l’ingénieur*, E1020 V1 (2004).

[99] *CODATA Internationally recommended 2018 values of the Fundamental Physical Constants*.

[100] Richard Feynman, Matthew Sands, and Robert B. Leighton. *The Feynman Lectures on Physics Vol. I Ch. 1: Polarization*. Caltech. URL: https://www.feynmanlectures.caltech.edu/I_33.html (visited on 02/19/2021).

[101] *Dictionnaire de l’Académie française. Dichroïsme*. Académie française. URL: https://www.dictionnaire-academie.fr/article/A9D2405 (visited on 02/24/2021).

[102] M Hamon et al. *Méthodes spectrales et analyse organique*. 2nd ed. Vol. 3. Chimie Analytique. Paris: Masson, 1998. ISBN: 9782225835070.

[103] Gregory Roos and Cathryn Roos. “Isomers and Stereochemistry”. In: *Organic Chemistry Concepts*. Elsevier, 2015, pp. 43–54. ISBN: 9780128016992. DOI: 10.1016/B978-0-12-801699-2.00003-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780128016992000031.

[104] *Circular Dichroism Unit Conversion. 4.Ultraviolet/visible spectroscopy*. Biozentrum Biophysics Facility, 2012. URL: https://www.biozentrum.unibas.ch/fileadmin/redaktion/05_Facilities/01_Technology_Platforms/BF/Protocols/CD_Unit_Conversion.pdf.

[105] *Circular Dichroism Spectroscopy Overview*. JASCO Inc. URL: https://jascoinc.com/learning-center/theory/spectroscopy/circular-dichroism-spectroscopy/ (visited on 03/25/2021).

[106] Alison Rodger. “Circular Dichroism Spectroscopy: Units”. In: *Encyclopedia of Biophysics*. Ed. by Gordon C. K. Roberts. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 316–317. ISBN: 9783642167119. DOI: 10.1007/978-3-642-16712-6_647. URL: http://link.springer.com/10.1007/978-3-642-16712-6_647.

[107] *MODEL J-715 Spectrophotometer. Hardware Manual*. P/N:0302-0265A. JASCO Corporation, 1995.

[108] Roberto Corradini et al. “Chirality as a tool in nucleic acid recognition: Principles and relevance in biotechnology and in medicinal chemistry”. en. In: *Chirality* 19.4 (2007), pp. 269–294. DOI: 10.1002/chir.20372.

[109] Walter A. Baase and W.Curtis Johnson. “Circular dichroism and DNA secondary structure”. In: *Nucleic Acids Research* 6.2 (1979), pp. 797–814. DOI: 10.1093/nar/6.2.797.

[110] Cindy A. Sprecher, Walter A. Baase, and W. Curtis Johnson. “Conformation and circular dichroism of DNA”. In: *Biopolymers* 18.4 (1979), pp. 1009–1019. DOI: 10.1002/bip.1979.360180418.

[111] Tomoo Miyahara, Hiroshi Nakatsuji, and Hiroshi Sugiyama. “Helical Structure and Circular Dichroism Spectra of DNA: A Theoretical Study”. en. In: *The Journal of Physical Chemistry A* 117.1 (2013), pp. 42–55. DOI: 10.1021/jp3085556.

[112] Wan Jun Chung et al. “Structure of a left-handed DNA G-quadruplex”. In: *Proceedings of the National Academy of Sciences* 112.9 (2015), pp. 2729–2733. DOI: 10.1073/pnas.1418718112.

[113] Boshi Fu et al. “Right-handed and left-handed G-quadruplexes have the same DNA sequence: distinct conformations induced by an organic small molecule and potassium”. In: *Chemical Communications* 52.65 (2016), pp. 10052–10055. DOI: 10.1039/C6CC04866H.

[114] Antonella Virgilio et al. “Unprecedented right- and left-handed quadruplex structures formed by heterochiral oligodeoxyribonucleotides”. In: *Biochimie* 93.7 (2011), pp. 1193–1196. DOI: 10.1016/j.biochi.2011.04.007.

[115] Fernaldo Richtia Winnerdy et al. “NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes”. In: *Nucleic Acids Research* 47.15 (2019), pp. 8272–8281. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz349.

[116] Wilma K Olson et al. “A standard reference frame for the description of nucleic acid base-pair geometry”. In: *Journal of Molecular Biology* 313.1 (2001), pp. 229–237. DOI: 10.1006/jmbi.2001.4987.

[117] Richard E. Dickerson. “Definitions and Nomenclature of Nucleic Acid Structure Parameters”. In: *Journal of Biomolecular Structure and Dynamics* 6.4 (1989), pp. 627–634. DOI: 10.1080/07391102.1989.10507726.

[118] R. Lavery et al. “Conformational analysis of nucleic acids revisited: Curves+”. In: *Nucleic Acids Research* 37.17 (2009), pp. 5917–5929. DOI: 10.1093/nar/gkp608.

[119] X.-J. Lu. “3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures”. In: *Nucleic Acids Research* 31.17 (2003), pp. 5108–5121. DOI: 10.1093/nar/gkg680.

[120] Vladimir Tsvetkov, Galina Pozmogova, and Anna Varizhuk. “The systematic approach to describing conformational rearrangements in G-quadruplexes”. In: *Journal of Biomolecular Structure and Dynamics* 34.4 (2016), pp. 705–715. DOI: 10.1080/07391102.2015.1055303.

[121] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD – Visual Molecular Dynamics”. In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.

[122] Roman V. Reshetnikov et al. “Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process”. In: *Nucleic Acids Research* 39.22 (2011), pp. 9789–9802. DOI: 10.1093/nar/gkr639.

[123] *Compendium of Chemical Terminology Gold Book*. Version 2.3.3. International Union of Pure and Applied Chemistry, 2014. URL: https://goldbook.iupac.org/files/pdf/goldbook.pdf.

[124] Tomoko Mashimo et al. “Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures”. In: *Journal of the American Chemical Society* 132.42 (2010), pp. 14910–14918. DOI: 10.1021/ja105806u.

[125] Horst Stöcker and Horst Stöcker. *Toutes les mathématiques et les bases de l’informatique*. OCLC: 1033430082. 2005. ISBN: 9782100702633.

[126] Roman Reshetnikov et al. “Structural Dynamics of Thrombin-Binding DNA Aptamer d(GGTTGGTGTGGTTGG) Quadruplex DNA Studied by Large-Scale Explicit Solvent Simulations”. In: *Journal of Chemical Theory and Computation* 6.10 (2010), pp. 3003–3014. DOI: 10.1021/ct100253m.

[127] R. V. Reshetnikov, A. V. Golovin, and A. M. Kopylov. “Comparison of models of thrombin-binding 15-mer DNA aptamer by molecular dynamics simulation”. In: *Biochemistry (Moscow)* 75.8 (2010), pp. 1017–1024. DOI: 10.1134/S0006297910080109.

[128] Gale Rhodes. *Crystallography made crystal clear: a guide for users of macromolecular models*. 3rd ed. Complementary science series. Amsterdam ; Boston: Elsevier/Academic Press, 2006. ISBN: 9780125870733.

[129] Bernadette Bensaude-Vincent and Richard-Emmanuel Eastes. *Philosophie de la chimie*. Collection Philosophie des sciences. Louvain-la-Neuve Paris: De Boeck supérieur, 2020. ISBN: 9782807305663.

[130] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: 10.1002/andp.19273892002.

[131] Andrew R. Leach. *Molecular modelling: principles and applications*. 2nd ed. Harlow, England ; New York: Prentice Hall, 2001. ISBN: 9780582382107.

[132] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. 2nd ed. Computational science series 1. San Diego: Academic Press, 2002. ISBN: 9780122673511.

[133] Christopher J. Cramer. *Essentials of computational chemistry: theories and models*. 2nd ed. Chichester, West Sussex, England ; Hoboken, NJ: Wiley, 2004. ISBN: 9780470091821.

[134] Horst Stöcker et al. *Toute la physique*. 2nd ed. Paris: Dunod, 2007. ISBN: 9782100511815.

[135] P. Fuchs. *Modélisation Moléculaire. Cours3: Méthodes dexploration de la surface dénergie potentielle*. Université Paris Diderot, 2007. URL: https://www.dsimb.inserm.fr/~fuchs/M1BI/Cours3_MD.pdf.

[136] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341. ISSN: 00219991. DOI: 10.1016/0021-9991(77)90098-5.

[137] Shuichi Miyamoto and Peter A. Kollman. "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962. ISSN: 1096-987X. DOI: 10.1002/jcc.540130805.

[138] Chad W. Hopkins et al. "Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning". In: *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1864–1874. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/ct5010406.

[139] Geoffrey M. Cooper. *The cell: a molecular approach*. 2. ed. Washington, DC: ASM Press, 2000. ISBN: 9780878931064.

[140] Massimiliano Bonomi and Carlo Camilloni, eds. *Biomolecular simulations: methods and protocols*. Methods in molecular biology 2022. New York, NY: Humana Press, 2019. ISBN: 9781493996070.

[141] Jeff Prucher, ed. *Brave new words: the Oxford dictionary of science fiction*. Oxford ; New York: Oxford University Press, 2007. ISBN: 9780195305678.

[142] Paul Arnaud. *Chimie générale: les cours de Paul Arnaud*. 8th ed. Paris: Dunod, 2016. ISBN: 9782100743674.

[143] Christophe Chipot. *Les methodes numériques de la dynamique moléculaire*. École de modélisation des macromolécules biologiques, 2002. URL: http://ecole.modelisation.free.fr/cours_chipot.pdf.

[144] Thérèse Malliavin. "Simulations de dynamique moléculaire en biochimie". In: *Techniques de l'ingénieur*, AF6043 V1 (2003).

[145] A. D. MacKerell et al. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins ". In: *The Journal of Physical Chemistry B* 102.18 (1998), pp. 3586–3616. DOI: 10.1021/jp973084f.

[146] David A. Case et al. "The Amber biomolecular simulation programs". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1668–1688. DOI: 10.1002/jcc.20290.

[147] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. "An overview of the Amber biomolecular simulation package: Amber biomolecular simulation package". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (2013), pp. 198–210. ISSN: 17590876. DOI: 10.1002/wcms.1121.

[148] Peter Kollman et al. "The development/application of a minimalist organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data". In: *Computer Simulation of Biomolecular Systems*. Ed. by Wilfred F. van Gunsteren, Paul K. Weiner, and Anthony J. Wilkinson. Dordrecht: Springer Netherlands, 1997, pp. 83–96. DOI: 10.1007/978-94-017-1120-3_2.

[149] Jay W. Ponder and David A. Case. "Force Fields for Protein Simulations". In: *Advances in Protein Chemistry*. Vol. 66. Elsevier, 2003, pp. 27–85. ISBN: 9780120342662. DOI: 10.1016/S0065-3233(03)66002-X.

[150] Wendy D. Cornell et al. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules”. In: *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197. DOI: 10.1021/ja00124a002.

[151] Thomas E. Cheatham and David A. Case. “Twenty-five years of nucleic acid simulations”. In: *Biopolymers* (2013), n/a–n/a. DOI: 10.1002/bip.22331.

[152] James A. Maier et al. “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB”. In: *Journal of Chemical Theory and Computation* 11.8 (2015), pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255.

[153] Chuan Tian et al. “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”. In: *Journal of Chemical Theory and Computation* 16.1 (2020), pp. 528–552. DOI: 10.1021/acs.jctc.9b00591.

[154] James B. Foresman and AEleen Frisch. *Exploring chemistry with electronic structure methods*. Third edition. Wallingford, CT: Gaussian, Inc, 2015. ISBN: 9781935522034.

[155] M. J. Frisch et al. *Gaussian16 Revision C.01*. Gaussian Inc. Wallingford CT. 2016.

[156] H. Bernhard Schlegel. “Optimization of equilibrium geometries and transition structures”. In: *Journal of Computational Chemistry* 3.2 (1982), pp. 214–218. DOI: 10.1002/jcc.540030212.

[157] H. Bernhard Schlegel. “Estimating the hessian for gradient-type geometry optimizations”. In: *Theoretica Chimica Acta* 66.5 (1984), pp. 333–340. ISSN: 0040-5744, 1432-2234. DOI: 10.1007/BF00554788.

[158] Chunyang Peng et al. “Using redundant internal coordinates to optimize equilibrium geometries and transition states”. In: *Journal of Computational Chemistry* 17.1 (1996), pp. 49–56. DOI: 10.1002/(SICI)1096-987X(19960115)17:1<49::AID-JCC5>3.0.CO;2-0.

[159] Christopher I. Bayly et al. “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model”. In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280. ISSN: 0022-3654, 1541-5740. DOI: 10.1021/j100142a004.

[160] Piotr Cieplak et al. “Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins”. In: *Journal of Computational Chemistry* 16.11 (1995), pp. 1357–1377. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.540161106.

[161] William Kemp. *Organic spectroscopy*. 3rd ed. Palgrave, 1991. ISBN: 9780333519547.

[162] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.202427399.

[163] Giovanni Bussi and Alessandro Laio. “Using metadynamics to explore complex free-energy landscapes”. In: *Nature Reviews Physics* 2.4 (2020), pp. 200–212. ISSN: 2522-5820. DOI: 10.1038/s42254-020-0153-0.

[164] Frank Noé and Cecilia Clementi. “Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods”. In: *Current Opinion in Structural Biology* 43 (2017), pp. 141–147. ISSN: 0959440X. DOI: 10.1016/j.sbi.2017.02.006. (Visited on 08/30/2022).

[165] Conrad Shyu and F. Marty Ytreberg. “Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data”. In: *Journal of Computational Chemistry* (2009), NA–NA. ISSN: 01928651, 1096987X. DOI: 10.1002/jcc.21231.

[166] Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. “Adaptive biasing force method for scalar and vector free energy calculations”. In: *The Journal of Chemical Physics* 128.14 (2008), p. 144120. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.2829861.

[167] Adrien Lesage et al. “Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method”. In: *The Journal of Physical Chemistry B* 121.15 (2017), pp. 3676–3685. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.6b10055.

[168] Haohao Fu et al. “Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys”. In: *The Journal of Physical Chemistry Letters* 9.16 (2018), pp. 4738–4745. ISSN: 1948-7185, 1948-7185. DOI: 10.1021/acs.jpclett.8b01994.

[169] G.M. Torrie and J.P. Valleau. "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199. ISSN: 00219991. DOI: 10.1016/0021-9991(77)90121-8.

[170] Johannes Kästner. "Umbrella sampling: Umbrella sampling". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942. ISSN: 17590876. DOI: 10.1002/wcms.66.

[171] Shankar Kumar et al. "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.540130812.

[172] Benoît Roux. "The calculation of the potential of mean force using computer simulations". In: *Computer Physics Communications* 91.1-3 (1995), pp. 275–282. ISSN: 00104655. DOI: 10.1016/0010-4655(95)00053-I. (Visited on 09/01/2022).

[173] Jean-Louis Basdevant. *15 leçons de mécanique quantique*. Louvain-la-Neuve (Belgique) Paris: De Boeck supérieur, 2019. ISBN: 9782807321786.

[174] Claude Leforestier. *Introduction à la chimie quantique*. Paris: Dunod, 2014. ISBN: 9782100081998.

[175] *The Nobel Prize in Chemistry 1998*. URL: https://www.nobelprize.org/prizes/chemistry/1998/8811-the-nobel-prize-in-chemistry-1998/ (visited on 09/03/2022).

[176] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4A (1965), A1133–A1138. ISSN: 0031-899X. DOI: 10.1103/PhysRev.140.A1133.

[177] A. Warshel and M. Levitt. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *Journal of Molecular Biology* 103.2 (), pp. 227–249. ISSN: 00222836. DOI: 10.1016/0022-2836(76)90311-9.

[178] *The Nobel Prize in Chemistry 2013*. URL: https://www.nobelprize.org/prizes/chemistry/2013/summary/.

[179] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[180] Buvaneswari Coimbatore Narayanan et al. “The Nucleic Acid Database: new features and capabilities”. In: *Nucleic Acids Research* 42.D1 (2014), pp. D114–D122. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkt980.

[181] Hugo Gattuso, Xavier Assfeld, and Antonio Monari. “Modeling DNA electronic circular dichroism by QM/MM methods and Frenkel Hamiltonian”. In: *Theoretical Chemistry Accounts* 134.3 (2015), p. 36. ISSN: 1432-881X, 1432-2234. DOI: 10.1007/s00214-015-1640-8.

[182] Ana Elisa Bergues-Pupo et al. “Role of the central cations in the mechanical unfolding of DNA and RNA G-quadruplexes”. In: *Nucleic Acids Research* 43.15 (2015), pp. 7638–7647. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv690.

[183] Martin Tomako, Michaela Vorlíková, and Janos Sagi. “Substitution of adenine for guanine in the quadruplex-forming human telomere DNA sequence G3(T2AG3)3”. In: *Biochimie* 91.2 (2009), pp. 171–179. ISSN: 03009084. DOI: 10.1016/j.biochi.2008.07.012.

[184] Brenna A. Tucker et al. “Stability of the Na^+ Form of the Human Telomeric G-Quadruplex: Role of Adenines in Stabilizing G-Quadruplex Structure”. In: *ACS Omega* 3.1 (2018), pp. 844–855. ISSN: 2470-1343, 2470-1343. DOI: 10.1021/acsomega.7b01649.

[185] Phillip A. Rachwal, Tom Brown, and Keith R. Fox. “Sequence effects of single base loops in intramolecular quadruplex DNA”. In: *FEBS Letters* 581.8 (2007), pp. 1657–1660. ISSN: 00145793. DOI: 10.1016/j.febslet.2007.03.040.

[186] Martina Lenari ivkovi, Jan Rozman, and Janez Plavec. “AdenineDriven Structural Switch from a Two to ThreeQuartet DNA GQuadruplex”. In: *Angewandte Chemie International Edition* 57.47 (2018), pp. 15395–15399. ISSN: 1433-7851, 1521-3773. DOI: 10.1002/anie.201809328. (Visited on 07/17/2022).

[187] Beatrice Karg and Klaus Weisz. “Loop Length Affects *Syn* *Anti* Conformational Rearrangements in Parallel GQuadruplexes”. In: *Chemistry A European Journal* 24.40 (2018), pp. 10246–10252. ISSN: 0947-6539, 1521-3765. DOI: 10.1002/chem.201801851.

[188] Aurore Guédin et al. “How long is too long? Effects of loop size on G-quadruplex stability”. In: *Nucleic Acids Research* 38.21 (2010), pp. 7858–7868. ISSN: 1362-4962, 0305-1048. DOI: 10.1093/nar/gkq639.

[189] Subramaniyam Ravichandran et al. “The effect of hairpin loop on the structure and gene expression activity of the long-loop G-quadruplex”. In: *Nucleic Acids Research* 49.18 (2021), pp. 10689–10706. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab739.

[190] Stas Bielskut, Janez Plavec, and Peter Podbevek. “Impact of Oxidative Lesions on the Human Telomeric G-Quadruplex”. In: *Journal of the American Chemical Society* 141.6 (2019), pp. 2594–2603. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.8b12748. (Visited on 07/19/2022).

[191] Natacha Driessens et al. “Hydrogen peroxide induces DNA single- and double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ”. In: *Endocrine-Related Cancer* 16.3 (2009), pp. 845–856. ISSN: 1351-0088, 1479-6821. DOI: 10.1677/ERC-09-0020.

[192] Cynthia J. Burrows and James G. Muller. “Oxidative Nucleobase Modifications Leading to Strand Scission”. In: *Chemical Reviews* 98.3 (1998), pp. 1109–1152. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr960421s.

[193] Lucia Lauková et al. “Deoxyribonucleases and Their Applications in Biomedicine”. In: *Biomolecules* 10.7 (2020), p. 1036. ISSN: 2218-273X. DOI: 10.3390/biom10071036.

[194] Keith W. Caldecott. “Single-strand break repair and genetic disease”. In: *Nature Reviews Genetics* 9.8 (2008), pp. 619–631. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2380.

[195] Anne-Sophie Banneville et al. “Structural and functional characterization of DdrC, a novel DNA damage-induced nucleoid associated protein involved in DNA compaction”. In: *Nucleic Acids Research* 50.13 (2022), pp. 7680–7696. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac563.

[196] Brad J. Schmier and Stewart Shuman. “Deinococcus radiodurans HD-Pnk, a Nucleic Acid End-Healing Enzyme, Abets Resistance to Killing by Ionizing Radiation and Mitomycin C”. In: *Journal of Bacteriology* 200.17 (2018). Ed. by William W. Metcalf. ISSN: 0021-9193, 1098-5530. DOI: 10.1128/JB.00151-18.

[197] Nitu Kumari et al. “G-quadruplex Structures Contribute to Differential Radiosensitivity of the Human Genome”. In: *iScience* 21 (2019), pp. 288–307. ISSN: 25890042. DOI: 10.1016/j.isci.2019.10.033. (Visited on 07/24/2022).

[198] Jinzhi Tan et al. “The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrodomains That Bind G-Quadruplexes”. In: *PLoS Pathogens* 5.5 (2009). Ed. by Félix A. Rey, e1000428. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1000428.

[199] Margaret A. Johnson et al. “SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding”. In: *Journal of Molecular Biology* 400.4 (2010), pp. 724–742. ISSN: 00222836. DOI: 10.1016/j.jmb.2010.05.027.

[200] V. Dhamodharan and P. I. Pradeepkumar. “Specific Recognition of Promoter G-Quadruplex DNAs by Small Molecule Ligands and Light-up Probes”. In: *ACS Chemical Biology* (2019), acschembio.9b00475. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/acschembio.9b00475.

[201] Debasish Dutta et al. “Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex”. In: *Nucleic Acids Research* 46.11 (2018), pp. 5355–5365. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky385.

[202] Joana Figueiredo et al. “Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs”. In: *Molecules* 26.20 (2021), p. 6164. ISSN: 1420-3049. DOI: 10.3390/molecules26206164.

[203] Saniya M Javadekar et al. “Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms”. In: *DNA Research* 27.5 (2020), dsaa024. ISSN: 1340-2838, 1756-1663. DOI: 10.1093/dnares/dsaa024. (Visited on 07/26/2022).

[204] Oliver Scholz, Simon Hansen, and Andreas Plückthun. “G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins”. In: *Nucleic Acids Research* 42.14 (2014), pp. 9182–9194. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gku571. (Visited on 07/26/2022).

[205] Andreas Plückthun. “Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy”. In: *Annual Review of Pharmacology and Toxicology* 55.1 (2015), pp. 489–511. ISSN: 0362-1642, 1545-4304. DOI: 10.1146/annurev-pharmtox-010611-134654. (Visited on 07/26/2022).

[206] Peer RE Mittl, Patrick Ernst, and Andreas Plückthun. "Chaperone-assisted structure elucidation with DARPins". In: *Current Opinion in Structural Biology* 60 (2020), pp. 93–100. ISSN: 0959440X. DOI: 10.1016/j.sbi.2019.12.009.

[207] Robert C Münch et al. "DARPins: An Efficient Targeting Domain for Lentiviral Vectors". In: *Molecular Therapy* 19.4 (2011), pp. 686–693. ISSN: 15250016. DOI: 10.1038/mt.2010.298.

Abstract in french

Les acides nucléiques sont des macromolécules organiques qui résultent de la polymérisation de nucléotides. Ces molécules sont généralement considérées comme le support de l'information génétique. Deux familles d'acides nucléiques sont actuellement connues : l'ADN et l'ARN. D'un point de vue structurel, la forme la plus populaire est la double hélice d'ADN. Cependant, d'autres formes existent et parmi elles, le G-quadruplex. Il s'agit d'un repliement de l'ADN, ou de l'ARN, dans une zone riche en guanines. Celles-ci forment des quadruplex de guanines, qui sont empilées les unes sur les autres et sont stabilisées par un cation central. Les structures G-quadruplex sont de plus en plus étudiées. Ceci n'est pas surprenant puisque leur rôle biologique implique la régulation de mécanismes génétiques. Ils sont notamment impliqués dans la régulation du cycle cellulaire, mais ils jouent également un rôle dans le cancer, certaines maladies neurologiques ou virales. L'objectif de cette thèse est d'étudier les G-quadruplex en utilisant les outils de la chimie théorique. Les trois années de travail soulèvent des points très importants pour la recherche sur les G-quadruplex. Tout d'abord, la modélisation d'une structure théorique de G-quadruplex peut être réalisée par homologie de séquence et validée par les calculs d'un spectre théorique de dichroïsme circulaire. Par conséquent, il est possible d'utiliser ces outils pour proposer et utiliser une structure G-quadruplex si elle n'est pas encore résolue expérimentalement. Ensuite, le travail effectué montre que les G-quadruplex forment un repliement très stable puisqu'ils sont globalement conservés même lorsque de la 8-oxo-guanine ou des lésions de rupture de brin sont introduites au niveau des quartets. Ensuite, l'article se concentre sur l'interaction entre les G-quadruplex et les protéines. Il met en évidence le rôle important de l'ARN G-quadruplex dans l'infection du pathogène viral SARS-CoV-2. Cet ARN favorise la dimérisation de la protéine SUD du virus, qui est à son tour responsable de la perturbation du système immunitaire. Enfin, cette thèse fournit une explication structurelle de l'interaction spécifique entre la protéine DARPin 2E4 et le G-quadruplex du promoteur c-Myc.

Modeling the influence of DNA lesion on the regulation of gene expression

Tom Miclot

Nucleic acids are organic macromolecules that result from the polymerization of nucleotides. These molecules are generally considered as the support of the genetic information. Two families of nucleic acids are currently known: DNA and RNA. From a structural point of view, the most popular form is the double helix of DNA. However, other forms exist and among them are the G-quadruplex. This is a folding of the DNA, or RNA, in an area rich in guanines. These form quadruplex of guanines, which are stacked on top of each other and are stabilized by a central cation. G-quadruplex structures are increasingly studied. This is not surprising since their biological role involves the regulation of genetic mechanisms. They are notably involved in the regulation of the cell cycle, but they also play a role in cancer, certain neurological or viral diseases. The aim of this PhD thesis is to study G-quadruplex using theoretical chemistry tools. The three years of work raise very important points for the research on G-quadruplex. First, the modeling of a theoretical G-quadruplex structure can be achieved by sequence homology and validated by calculations of a theoretical circular dichroism spectrum. Consequently, it is possible to use these tools to propose and use a G-quadruplex structure if it is not yet experimentally solved. Then, the work done shows that G-quadruplex form a very stable folding since they are globally conserved even when 8-oxo-guanine or strand breaks lesions are introduced at the quartets. Then, the paper focuses on the interaction between G-quadruplex and proteins. It highlights the important role of G-quadruplex RNA in the infection of the viral pathogen SARS-CoV-2. This RNA promotes the dimerization of the SUD protein of the virus, which in turn is responsible for the disruption of the immune system. Finally, this thesis provides a structural explanation for the specific interaction between the DARPin 2E4 protein and the G-quadruplex of the c-Myc promoter.

Keywords G-quadruplex, Interaction DNA-protein, Damaged DNA, Molecular dynamic